IDENTIFICATION OF A SUITABLE SNP FOR ALLELE-SPECIFIC SILENCING OF THE DISEASE-CAUSING GENE IN SCA1 PATIENTS IN SOUTH AFRICA

FIONA KEBIRUNGI BAINE

BNXFIO003

Submitted to the University of Cape Town

In fulfilment of the requirements for the degree of MSc (MED) in Human Genetics

August 2010

CHAPTER ONE:

INTRODUCTION

CHAPTER TWO:

MATERIALS AND METHODS

CHAPTER THREE:

RESULTS

CHAPTER FOUR:

DISCUSSION

CHAPTER FIVE:

FUTURE PROSPECTS AND CONCLUSION

APPENDICES

REFERENCES

Declaration

I, Fiona Kebirungi Baine, hereby declare that the work on which this research project is based is my own original work (except where acknowledgements indicate otherwise), and that neither the whole work or any part thereof has been, is being, or will be submitted for another degree at this or any other university. I empower the university to reproduce for the purpose of research either the whole, or any portion of the contents in any manner whatsoever.

This study has been approved by the Research Ethics Committee of the Faculty of Health Sciences, UCT; reference number 253/2008.

Name: Fiona Kebirungi Baine

Student number: BNXFIO003

Signature:

Date:

Citations and references are made according to recommendations to authors from the European Journal of Human Genetics.

Table of Contents

2. Materials and Methods

SECTION I: Construction of a SNP-based haplotype around the SCA1 locus

SECTION II: Allele-specific silencing

3. Results

SECTION I: Construction of a SNP-based haplotype around the SCA1 locus

SECTION II: Allele-specific silencing

4. Discussion

5. Future prospects and conclusion

This work is dedicated to my grandfathers:

TIBAMANYA MWENE MUSHANGA

and

LAZARUS HOKUKIRIRE (R.I.P)

"THANK YOU FOR STARTING THE BALL ROLLING"

ACKNOWLEDGEMENTS

To my supervisors and co-supervisors: Professors LJ Greenberg and A Bryer and Doctors M Wood and J Scholefield – thank you for all your help, encouragement and hard-work in getting this project to completion.

To Dr M Weinberg and the members of the Antiviral Gene Therapy Research Unit (AGTRU), University of the Witswatersrand – thank you for your hospitality and assistance with the techniques of cloning, transfection and gene expression assays.

To Ms I Baumgarten and Mr B Allan – thank you for all of your assistance with the techniques of cell culture.

To the entire Human Genetics Division – you are and have always been wonderful! Thank you for all the discussions, the laughs and for being family.

To my friends in my two homes and everywhere else – your love and friendship is greatly treasured and I will always love you.

To my parents, my sisters and brother who have loved encouraged and supported me in every way imaginable – I love you enormously and I'm eternally grateful.

To Ticha, my best friend, my love – thank you for everything big and small and especially for keeping me sane!

List of figures

List of tables

List of abbreviations

3’	3 prime
5’	5 prime
17-AAG	17-(allylamino)-17-demethoxygeldanamycin
°C	degrees Celsius
µl	microlitre(s)
µg	microgram(s)
A	adenine
AAV	adeno-associated virus
Ad	adenovirus
AGO2	Argonaute 2
Akt	v-akt murine thymoma viral oncogene homolog 1
APS	ammonium persulphate
ATP	adenosine triphosphate
ATXN1	ataxin 1 (human gene)
Atxn1L	ataxin 1-like (mouse gene)
ATXN3	ataxin 3 (human gene)
ATXN7	ataxin 7 (human gene)
AXH	ataxin 1/High-mobility group box 1
BLAST	basic local alignment search tool
Boat	Brother of ataxin-1
bp	base pairs
C	cytosine
C-	carboxyl
C. elegans	*Caenorhabditis elegans*
$CaCl_2$	calcium chloride
CAG	DNA sequence encoding glutamine
cDNA	complementary DNA

CIC/Cic	Capicua
CNS	central nervous system
D. melanogaster	*Drosophila melanogaster*
DGCR8	Di-George syndrome critical region 8
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotide triphosphates
DRPLA	Dentatorubralpallidoluysian atrophy
dsDNA	double stranded DNA
dsRed	red fluorescent protein
dsRNA	double-stranded RNA
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetra-acetic acid
EtBr	ethidium bromide
eGFP	expressed green fluorescent protein
F	forward
g	gram
G	guanine
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
HD	Huntington disease
HEK	human embryonic kidney
HWE	Hardy-Weinberg equilibrium
Hsp/HSP	Heat shock protein
kb	kilobase pairs
kDa	kiloDalton (s)
*lac*Z	β-galactosidase
LANP	leucine-rich acidic nuclear protein
LB	Luria broth
LE	low electroendosmosis

LN	large normal
LV	lentivirus
M	molar
MAF	minor allele frequency
Mb	megabase pairs
mg	milligram(s)
$MgCl_2$	magnesium chloride
min	minute(s)
miRNA	micro RNA
ml	millilitre(s)
mM	millimole(s)
mRNA	messenger RNA
NCBI	National Centre for Biotechnology Information
ng	nanogram(s)
NIIs	neuronal intranuclear inclusions
NLS	nuclear localisation signal
nm	nanometres
nt	nucleotide(s)
P-bodies	processing bodies
PBS	phosphate buffered saline
Pcp2	Purkinje cell protein 2
PCR	polymerase chain reaction
PDA	piperazine diacrylamide
pDNA	plasmid DNA
pg	picogram(s)
PIWI	P-element induced wimpy testis
pmol	picomoles
Pol	polymerase

polyQ	polyglutamine
PQB1	polyQ-binding protein
pre-miRNA	precursor micro RNA
pri-miRNA	primary micro RNA
R	reverse
RBM17	RNA-binding motif protein 17
RE	restriction endonuclease
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
rpm	revolutions per minute
s	second(s)
SBMA	spinobulbar muscular atrophy
Sca1	spinocerebellar ataxia 1 (mouse gene)
SCA(s)	spinocerebellar ataxia(s)
SDS	sodium dodecyl sulfate
Ser776	serine residue at position 776
sH_2O	sterile water
shRNA	short hairpin RNA
siRNA	short interfering RNA
SMRT	silencing mediator of retinoid and thyroid hormone receptors
SNP	single nucleotide polymorphism
Sp-1	specificity protein 1
SSCP	single-strand conformational polymorphism
T	thymine
T_a	annealing temperature
T_m	melting temperature
TBE	tris borate EDTA buffer

TEMED	Tetramethylethylenediamine
U	uracil
UCT	University of Cape Town
UHM	U2A homology motif
ULM	UHM ligand motif
UTR	untranslated region
UV	ultraviolet
v	version
V	volts
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
XPO5	exportin 5

ABSTRACT

Spinocerebellar ataxia 1 (SCA1) is part of a broader group of dominant neurodegenerative disorders caused by an unstable CAG trinucleotide repeat. There is no known cure for this disease and symptoms worsen progressively culminating in death. The disease-causing mutation in SCA1 occurs in the *ATXN1* gene. The function of the gene product (the ataxin-1 protein) is unknown, however; the protein has been linked to RNA processing in the cell.

The first part of this study followed on a 1997 report of two founder haplotypes in the Mixed Ancestry SCA1 families in the Western Cape of South Africa, using microsatellites. The aim was to narrow the region investigated in the previous study, and confirm the existence of founder haplotypes using a SNP-based haplotype. The SNP-based haplotype was constructed using 4 SNPs in individuals from 5 different families of Mixed Ancestry origin from the Western Cape and the two founder events were confirmed. The SNP-based haplotype also shows the existence of a minimum common interval and indicates regions of possible break-points which may be useful in determining the extent and origins of the two haplotypes.

The second aim of the study was to preferentially silence the mutant transcript of the *ATXN1* gene by targeting a single nucleotide difference. Two of the SNPs genotyped for the SNP-based haplotype were found to be heterozygous in over half of the patient cohort. Eight shRNA effector molecules were screened against short target sequences incorporating one of these SNPs. Results are promising, with significant discrimination achieved between the wild-type and mutant alleles by targeting this SNP. This study has shown that RNAi may be developed as a beneficial therapeutic technique for a subset of SCA1 patients in South Africa.

1. Introduction

1.1 Polyglutamine disorders

The polyglutamine (polyQ) diseases are a large group of inherited neurodegenerative disorders. They include Huntington disease (HD), dentatorubral-pallidoluysian atrophy (DRPLA), spinobulbar muscular atrophy (SBMA) as well as the spinocerebellar ataxias type 1, 2, 3, 6, 7 and 17 (Zoghbi and Orr 2000, Soong and Paulson 2007). For these nine known polyQ disorders, the disease-causing mutation is an unstable expansion of a trinucleotide (CAG) repeat in the respective gene that leads to an abnormally long polyglutamine tract in the protein. Transmission of these large unstable alleles results in further expansion in subsequent generations, a phenomenon known as 'anticipation' (Margolis 2002, Soong and Paulson 2007). Apart from SBMA, the polyQ diseases are dominantly inherited, and longer CAG tracts usually correlate with an earlier age-of-onset as well as more severe symptoms.

The genes causing these degenerative disorders show no homology to each other outside of the glutamine repeat, and most encode proteins that are ubiquitously expressed (Zoghbi and Orr 2000). Despite widespread expression of the gene product, each disease is characterised by a unique pattern of selective neuronal death within the central nervous system (CNS) that leads to the observed clinical symptoms. The exact mechanism for this selectivity remains elusive. The dominant pattern of inheritance as well as evidence from animal models supports a model of disease in which polyglutamine expansion confers a toxic gain-of-function on the protein (Ross 1997, Zoghbi and Orr 2000, Chen *et al.* 2003).

The neurotoxic properties probably involve an increased propensity for the disease protein to misfold and aggregate, resulting in the formation of neuronal intranuclear inclusions (NIIs) which have been found in neurons susceptible to the disease process in the polyQs (Skinner *et al.* 1997, Perez *et al.* 1998, Yamada *et al.* 2008).

1.2 Spinocerebellar ataxia 1

Spinocerebellar ataxia 1 (SCA1) is one of a group of autosomal dominant neurodegenerative disorders known as the spinocerebellar ataxias (SCAs). SCA1 is reported to account for 6% of all dominant ataxia (La Spada and Taylor 2010). Like all the inherited ataxias, SCA1 is characterised by loss of motor coordination and balance. More specific clinical features include slurred speech, difficulty swallowing, spasticity and cognitive impairments (Zoghbi and Orr 2009). In 1993, Orr and colleagues identified the disease-causing mutation in SCA1, showing it to be an expansion of a trinucleotide repeat (CAG) in the *ATXN1* gene which encodes the ataxin-1 protein (Orr *et al.* 1993). It was the first of the SCAs for which the causative gene was identified and characterised. The mutation leads to an abnormally long polyglutamine tract in the translated protein, the distinguishing feature of polyQ disorders.

In SCA1, selective neuronal death of the cerebellar Purkinje cells, inferior olivary cells and other brainstem nuclei, results in the observed symptoms (Klement *et al.* 1998). The disease is progressive and as yet incurable. Symptoms typically appear between 20 and 50 years of age and worsen progressively; resulting in death 10 to 20 years after the initial onset (Burright *et al.* 1995, Bryer *et al.* 2003). There are also reports of juvenile onset SCA1, where disease symptoms are more severe and progression is more rapid than the adult-onset form (Cummings *et al.* 1999).

Normal SCA1 alleles are highly polymorphic with 6 to 39 CAG repeats (Zühlke *et al.* 2002). The alleles in the upper limit of this range (36-39) are known as mutable normal or intermediate alleles. An analysis of these intermediate alleles performed by Zühlke and colleagues suggests a change from normal to pathogenic at 39 repeats depending on the presence of CAT interruptions within the CAG tract (Zühlke *et al.* 2002). Normal alleles with more than 20 repeats usually have between 1 and 3 CAT interruptions which are believed to be important in maintaining their stability during DNA replication, making them unsusceptible to intergenerational expansion (Chung *et al.* 1993, Orr and Zoghbi 2001). Pure CAG tracts however are dynamic; thus normal alleles carrying a pure tract are likely to expand into the disease-causing range on transmission. Symptomatic individuals have been reported to carry expanded alleles with 39 to 81 uninterrupted CAG repeats. An individual with juvenile-onset SCA1 was found to carry 91 repeats (Chong *et al.* 1995). Due to the unstable nature of the pure CAG tract, expanded alleles from symptomatic individuals may result in anticipation in the next generation (Orr and Zoghbi 2001, Zühlke *et al.* 2002, Soong and Paulson 2007). Anticipation is a common phenomenon in SCA1 and refers to the tendency for a disease to worsen from one generation to the next within the same family as a result of larger expansions causing the earlier onset of disease symptoms (Soong and Paulson 2007).

Chong *et al.* reported somatic instability of the *ATXN1* CAG tract especially in tissues with higher mitotic potential such as sperm cells and peripheral blood cells (Chong *et al.* 1995). This may explain the observation of an unbalanced transmission of expanded alleles with larger alleles being transmitted by the male parent (Chung *et al.* 1993, Jodice *et al.* 1994). In contrast, maternal transmission of expanded alleles tends to be stable and often yields contraction in repeat number (Jodice *et al.* 1994).

1.3 *ATXN1*

The *ATXN1* gene on the short arm of chromosome 6 consists of nine exons, spanning approximately 450 kilobase pairs (kb) of genomic DNA. The first seven exons cover the 5' untranslated region (UTR) and the last two large exons 8 and 9, cover the coding region and the 3' UTR. The coding region is only 2448 base pairs (bp) long (Zoghbi *et al.* 1991, Banfi *et al.* 1994, Servadio *et al.* 1995). Importantly, *ATXN1* undergoes alternative splicing in various tissues. The gene encodes ataxin-1, a 98 kiloDalton (kDa) soluble protein of approximately 816 amino acids (aa) depending on the number of CAG repeats. The protein is expressed in a variety of cells of the nervous system as well as in peripheral tissues; however its specific function is as yet unknown. Immunohistochemical studies have revealed that expression of endogenous ataxin-1 is cytoplasmic in non-neuronal cell types and nuclear in neuronal cells of the brain (Servadio *et al.* 1995, Burright *et al.* 1995). However, in the Purkinje cells of the cerebellum the site of SCA1 pathology, endogenous ataxin-1 has both nuclear and cytoplasmic localization.

The protein consists of two functionally different domains: the ataxin-1/HBP1 (High-mobility group box 1) – AXH domain near the C-terminus of the protein and the amino terminal region which contains the polyglutamine tract. The protein also has a C-terminal nuclear localisation signal (NLS) that serves to target it to the nucleus (Klement *et al.* 1998). The AXH domain has been implicated in self-association (Burright *et al.* 1997), RNA-binding (Yue *et al.* 2001) and it is known to facilitate ataxin-1's interaction with several proteins (Chen *et al.* 2004, Tsuda *et al.* 2005, Lam *et al.* 2006).

Ataxin-1's ability to bind to RNA as well as its interactions with numerous proteins, most of which are involved in transcriptional regulation and RNA metabolism, point to a possible role in RNA processing. The protein has been shown to interact with several transcriptional co-regulators including: polyQ-binding protein 1 (PQB1) (Okazawa *et al.* 2002), silencing mediator of retinoid and thyroid hormone receptors (SMRT) (Tsai *et al.* 2004), Gfi-1/Senseless (Tsuda *et al.* 2005), leucine-rich acidic nuclear protein (LANP) (Cvetanovic *et al.* 2007), specificity protein 1 (Sp-1) (Goold *et al.* 2007) and the ubiquitin-conjugating enzyme UbcH6 (Lee *et al.* 2008).

A recent study explored the very complex process of ubiquitination and the role ataxin-1 plays within that process (Hong *et al.* 2008). The ubiquitin system-mediated post-translational modification and degradation of proteins is important for regulating the activity of normal proteins and to avoid the toxic effect of misfolded proteins. Thus, ataxin-1 appears to have an important role in cellular homeostasis.

1.4 SCA1 pathogenesis

The 1990's have been referred to as the "decade of the brain" (Cummings *et al.* 1999). This period saw rapid progress in neurogenetics research and on the SCAs in particular.

Early work attempted to elucidate the normal function of ataxin-1. Burright *et al.* established the first mouse model for SCA1 (Burright *et al.* 1995). The group generated transgenic mice expressing the human SCA1 gene using the promoter elements of the murine *Pcp2* gene that directs expression specifically to the Purkinje cells. The transgenes contained either a normal interrupted allele with 30 repeats or an expanded uninterrupted allele with 82 repeats.

The animals with the expanded SCA1 allele developed ataxia and loss of Purkinje cells (Burright *et al.* 1995). On the other hand, the transgenic mice expressing the unexpanded human SCA1 allele exhibited normal Purkinje cell numbers indicating that expression of expanded CAG repeats was sufficient to cause degeneration and ataxia.

To further investigate the function of ataxin-1, another mouse model was generated with a targeted deletion in the mouse *Sca1* gene (Matilla *et al.* 1998). Mice lacking ataxin-1 did not show evidence of ataxia or neuronal cell loss. However, *Sca1* null mice did show neurobehavioral abnormalities associated with learning and memory. These findings pointed to a gain-of-function of the mutant protein and a possible role of ataxin-1 in learning and memory (Matilla *et al.* 1998).

1.4.1 Protein aggregation

At this stage, nuclear aggregates containing the disease proteins had been discovered in a number of polyQ diseases including SCA1 (Skinner *et al.* 1997) and HD (Davies *et al.* 1997). The aggregates were identified post-mortem in brain tissue from human patients as well as in neurons taken from transgenic mice. Aggregation was therefore suggested as the common mechanism of neuropathology (Ross 1997, Perez *et al.* 1998, Davies *et al.* 1998). To examine this hypothesis in SCA1, Klement and colleagues (Klement *et al.* 1998) generated two lines of transgenic mice both with an expanded form of ataxin-1, one containing a mutated NLS and the other a deletion in the previously identified self-association region (Burright *et al.* 1997). In the first line (carrying the mutated NLS), the mutant form of the protein localised largely to the cytoplasm and the mice did not develop ataxia or degeneration (Klement *et al.* 1998). The second line (carrying the deletion) in which the mutant protein was able to localise to the nucleus showed progressive Purkinje cell pathology and neurological symptoms.

Therefore, although nuclear localisation is critical, nuclear aggregation of ataxin-1 is not necessary for pathogenesis. Aggregates have also been shown not to correlate with localisation of cellular protein in other polyQ diseases such as HD (Klement *et al.* 1998).

Several other studies have provided further understanding of the pathogenic mechanism underlying the polyQ disorders using SCA1 as a basis. A study of knock-in mice with 78 CAG repeats in the mouse *Sca1* locus showed no evidence of neuropathological changes up to 18 months of age (Lorenzetti *et al.* 2000). The mice homozygous for mutant ataxin-1 showed impaired motor coordination and intergenerational repeat instability compared to their wild-type littermates. However, histological examination of brain tissue from mutant mice up to 18 months showed none of the changes observed in other models over expressing expanded polyglutamine tracts. This data suggested that prolonged exposure to mutant ataxin-1 is necessary to produce the SCA1 phenotype and pathogenesis is thus a function of both the polyglutamine expanded protein and the duration of neuronal exposure to mutant protein (Lorenzetti *et al.* 2000). This could explain the late onset of disease symptoms seen in the majority of SCA1 patients.

The formation of NIIs (mentioned in section 1.1), although not required for pathogenesis was still thought to play an important role (Kaytor and Warren 1999). NIIs have been shown to contain ubiquitin and molecular chaperones, components of the cell's protein degradation and re-folding machinery. Cummings *et al.* hypothesised that protein misfolding and impaired clearance may lead to observed SCA1 pathology (Cummings *et al.* 1999). SCA1 transgenic mice crossed with mice over-expressing the molecular chaperone, inducible heat shock protein 70 (iHSP70), demonstrated that high levels of HSP70 protect against neurodegeneration, suggesting the upregulation of molecular chaperones as a possible therapeutic strategy (Cummings *et al.* 2001).

Several reports have shown the benefit of this strategy in other polyQ disorders. Geranylgeranylacetone, an acyclic isoprenoid compound used as an oral anti-cancer drug was previously shown to induce expression of heat shock proteins (Hsps) in different tissues. Oral administration of this compound led to improved motor behaviour and survival of transgenic SBMA mice (Bauer and Nukina 2009). Another compound, 17-(allylamino)-17-demethoxygeldanamycin (17-AAG) a known Hsp90 inhibitor, has also been reported to improve neurodegeneration in SBMA mice. In *Drosophila* models of SCA3 and HD, 17-AAG has been shown to suppress polyglutamine-induced neurodegeneration by activation of heat-shock factor 1 which in turn up-regulated molecular chaperones Hsp40, Hsp70 and Hsp90. In another study, an analogue of 17-AAG was used to ameliorate the phenotype in SBMA transgenic mice (Tokui *et al.* 2009). The use of molecular chaperones in SCA1 therapy thus requires further investigation.

1.4.2 SCA1 selectivity

Different hypotheses have been put forward to explain the selective nature of neuronal cell death. One hypothesis was that cell-specific expression levels and patterns of other proteins such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and leucine-rich acidic nuclear protein (LANP) interacting with the polyglutamine expanded ataxin-1, contributed to the observed selectivity (Timchenko and Caskey 1999). Another suggested that the ability to process and degrade toxic mutant proteins differed between neuronal cells, resulting in the degeneration of the more vulnerable cell types (Watase *et al.* 2002). Yet another hypothesis proposed somatic instability and further repeat expansion in particular cell types as a mechanism for the selectivity (Kennedy and Shelbourne 2000).

Although huntingtin protein expression levels are unchanged in HD knock-in mice, experiments showed unexpectedly high levels of repeat length variation in different tissues. The authors suggested that mechanisms of repeat instability in terminally differentiated tissue may differ significantly from those in the germ line and thus multiple rounds of DNA damage and repair could mediate mutation instability in post-mitotic cells (Kennedy and Shelbourne 2000).

An investigation by Watase and colleagues addressed the somatic and intergenerational repeat instability of the expanded allele in a knock-in SCA1 mouse model (Watase *et al.* 2003). The group studied maternal and paternal transmission of expanded alleles and the size of expansions in different somatic tissues. Results showed that the degree of instability was much less remarkable in the cerebellum compared to that in the striatum of mutant animals. This and the failure to find any 'hyper'-expanded allele in the cerebellum suggests the absence of a direct link between somatic CAG repeat instability and selective neuronal vulnerability (Watase *et al.* 2003). The reason for neuronal selectivity in the polyQs has still not been defined.

1.4.3 Post-translational modifications and protein interactions

In 2003, two groups reported the importance of post-translational modifications in the ataxin-1 protein. Emamian and colleagues showed that the serine at position 776 (S776) in both wild-type and mutant ataxin-1 was phosphorylated in transfected cells and transgenic mice. Changing the serine at this position to an alanine, suppressed the toxicity of the mutant protein. Phosphorylation was shown to be linked to the protein's transport into the nucleus and increased the protein's ability to form NIIs, both of which are essential for SCA1 pathogenesis (Emamian *et al.* 2003).

In another study using a *Drosophila* model of SCA1, Chen and colleagues reported the interaction of both wild-type and mutant ataxin-1 with the 14-3-3 protein, a multifunctional regulatory molecule (Chen *et al.* 2003). The authors suggest that 14-3-3 regulates ataxin-1's clearance under normal physiological conditions, either by stabilising the protein or impeding its interaction with other proteins that facilitate clearance. Expansion enhances this interaction and the mutant protein is further stabilised resulting in accumulation and aggregate formation. Chen *et al.* also reported that Akt is responsible for ataxin-1's phosphorylation at S776. Thus both 14-3-3 and Akt modulate neurodegeneration; Akt phosphorylates ataxin-1 promoting its binding to 14-3-3 which in turn leads to accumulation of ataxin-1 and neurodegeneration (Chen *et al.* 2003).

The interaction between ataxin-1 and 14-3-3 proteins has been suggested as a new potential target for therapeutic intervention in SCA1 (Umahara and Uchihara 2010). The 14-3-3 proteins have been shown to be directly involved in SCA1 pathogenesis with larger expansions having a higher affinity for 14-3-3 proteins. Ataxin-1 phosphorylation is required for the interaction, and 14-3-3 proteins have been found localised to NIIs in human SCA1 brain tissue. These highlighted factors point to 14-3-3 as an important interacting protein and a possible target for therapy (Umahara and Uchihara 2010).

Recently, de Chiara and colleagues reported the presence of a UHM ligand motif (ULM) in the C-terminus of ataxin-1 (de Chiara *et al.* 2009). This motif has been found in many splicing factors and its presence in ataxin-1 points to the protein's participation in a network of interactions (interactome) involving other splicing factors. An investigation into the mechanisms of ataxin-1's interactome found that the ULM overlaps both the NLS and the 14-3-3 binding motif.

This region, which also comprises S776, mediates ataxin-1's interaction with UHM domains of other proteins via the ULM. In its wild-type unexpanded form, un-phosphorylated ataxin-1 contributes to the spliceosome by interacting with various splicing factors via the ULM. The un-phosphorylated form of expanded ataxin-1 recognises UHM domains despite being unable to bind to 14-3-3. This could protect against self-association of the mutant protein allowing its clearance via the proteasome pathway. When phosphorylated, the expanded ataxin-1 binds to 14-3-3 altering its interactions with the spliceosome and other proteins. The authors suggest that some of these changes result in SCA1 disease pathology while other alterations, such as that mentioned above, may be protective (de Chiara *et al.* 2009).

Two important protein modification cellular processes, ubiquitination and SUMOylation have also been shown to be important in SCA1 pathogenesis. SUMOylation involves a group of proteins (small ubiquitin-like modifiers - SUMO), that are covalently attached to target proteins making them resistant to degradation. Riley *et al.* reported that the SUMOylation of ataxin-1 is dependent on the length of the polyglutamine tract, phosphorylation at Ser776 and nuclear localisation of the protein (Riley *et al.* 2005). Expansion of the tract resulted in a decrease in SUMOylation which may interfere with the role of ataxin-1 as a transcription regulator.

Ubiquitination is a process by which proteins are covalently tagged with ubiquitin making them recognisable to the cell's degradation machinery (Cummings *et al.* 1998). Ubiquitination involves 3 classes of enzymes, ubiquitin-activating enzyme E1, the ubiquitin-conjugating enzyme E2 and the ubiquitin-protein ligase E3 as shown in figure 1.1.

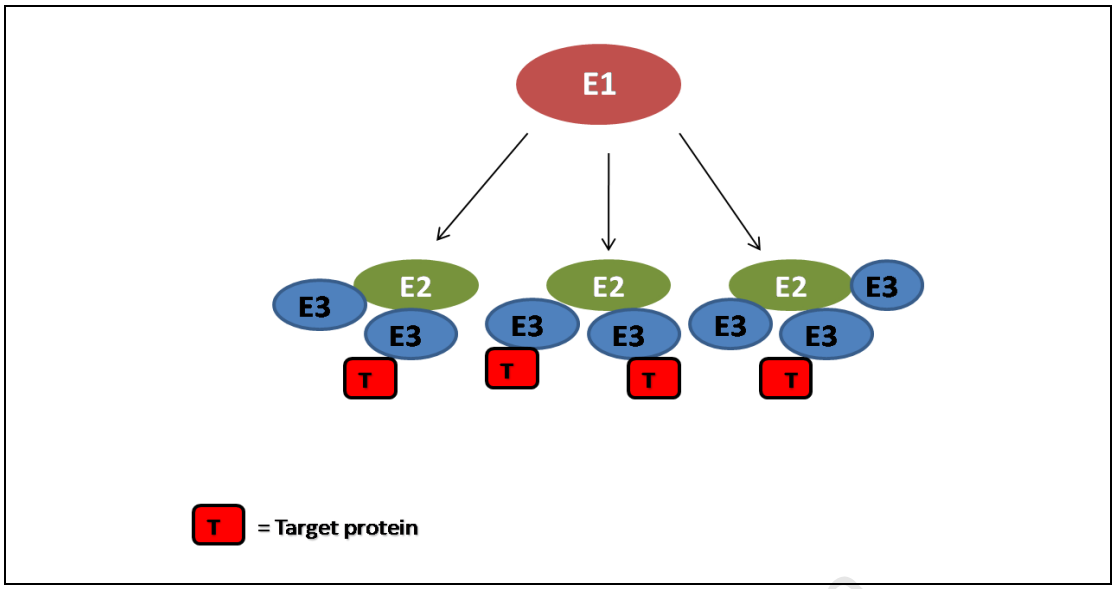


Figure 1.1 Pyramidal structure of enzyme interactions in the ubiquitination process. A single E1 activates many E2s which in turn interact with several E3s to tag proteins marked for degradation.

The enzyme classes (E1, E2 and E3) operate in a pyramidal structure with a single E1 activating many E2s which subsequently interact with several E3s (figure 1.1), thus selective degradation of proteins is mediated by E2s in conjunction with E3s (Hong *et al.* 2008). A study by Hong and colleagues showed that the E2 UbcH6 directly ubiquitinates wild-type ataxin-1, demonstrating that UbcH6 and ataxin-1 proteins are an E2-substrate pair in the ubiquitin-proteasome system. It is likely that the expanded form of the protein disrupts this interaction interfering with the degradation of mutant protein and leading to protein aggregation and toxicity (Hong *et al.* 2008).

A recently characterised protein, sacsin, has been reported to protect against polyglutamine-expanded ataxin-1 toxicity (Parfitt *el al.* 2009). The protein is widely expressed in the brain and at particularly high levels in the Purkinje cells, the site of SCA1 pathology. It was found to localise with the polyglutamine-expanded ataxin-1 inclusions, suggesting a possible role in the degradation of mutant protein.

Using bioinformatics, the protein was shown to contain a putative ubiquitin-like (UbL) domain that co-precipitated with the 20S proteasomal sub-unit, linking sacsin to the ubiquitin-proteasome system in which ataxin-1 is known play a role (Hong *et al.* 2008). The protein also contains a J-domain (a feature of DnaJ/Hsp40 proteins), which was shown to function with an *E.coli* Hsp70. This confirms sacsin as a type III Hsp40 protein. In addition to the domains discussed here, sacsin also contains several regions of homology to Hsp90, pointing to a possible chaperone function. It is however, not yet clear how sacsin's protective function is exerted (Parfitt *et al.* 2009).

In 1997, Skinner and colleagues reported that normal ataxin-1 localised to several nuclear structures while the expanded form localised to a single structure (Skinner *et al.* 1997). It has been recently reported that both wild-type and expanded ataxin-1 form soluble aggregates whereas other polyQ aggregates are insoluble structures (Krol *et al.* 2008). The group used Cos-7 cells transfected with different polyQ proteins fused to green fluorescent protein (GFP) to show these differences. Unexpectedly, while other polyglutamine-expanded proteins formed aggregates (static structures made up of tightly aggregated proteins), both wild-type and expanded ataxin-1 formed multiple nuclear accumulations. The ataxin-1 accumulations were found to be SDS-soluble unlike other polyQ proteins which were found to be SDS-insoluble. The nuclear accumulations of polyglutamine-expanded ataxin-1 were also observed to be more mobile and fused more frequently with each other compared to the unexpanded form of the protein (Krol *et al.* 2008). The increased dynamics may contribute to pathogenesis by further interfering with protein-protein interactions and the function of ataxin-1.

Ataxin-1 has been shown to associate with the nuclear matrix (Skinner *et al.* 1997). It is likely that the structures containing the expanded form of the protein are less stably associated with the matrix which explains the increased mobility (Krol *et al.* 2008). Ataxin-1 is also involved in several previously highlighted processes and pathways such as gene transcription and mRNA splicing, any of which may be affected by an increase in the nuclear dynamics of the expanded protein, thus contributing to SCA1 pathogenesis. Krol and colleagues suggest re-defining the observed ataxin-1 structures as accumulations rather than aggregates and highlight the importance of understanding the nature of these nuclear bodies in both the nucleus and cytoplasm (Krol *et al.* 2008).

In light of the fact that current data suggests SCA1 pathogenesis occurs via a protein- rather than an RNA-mediated mechanism, a model supporting altered protein interactions is more likely. Recent work on SCA1 has thus focused on identifying a pathogenic gain-of-function exerted by the expanded protein. One approach has been to identify proteins that interact with the disease protein under normal and pathological circumstances by elucidating the pathways in which ataxin-1 is involved. Figure 1.2 summarises the main avenues via which the mutant protein is believed to exert its effects.

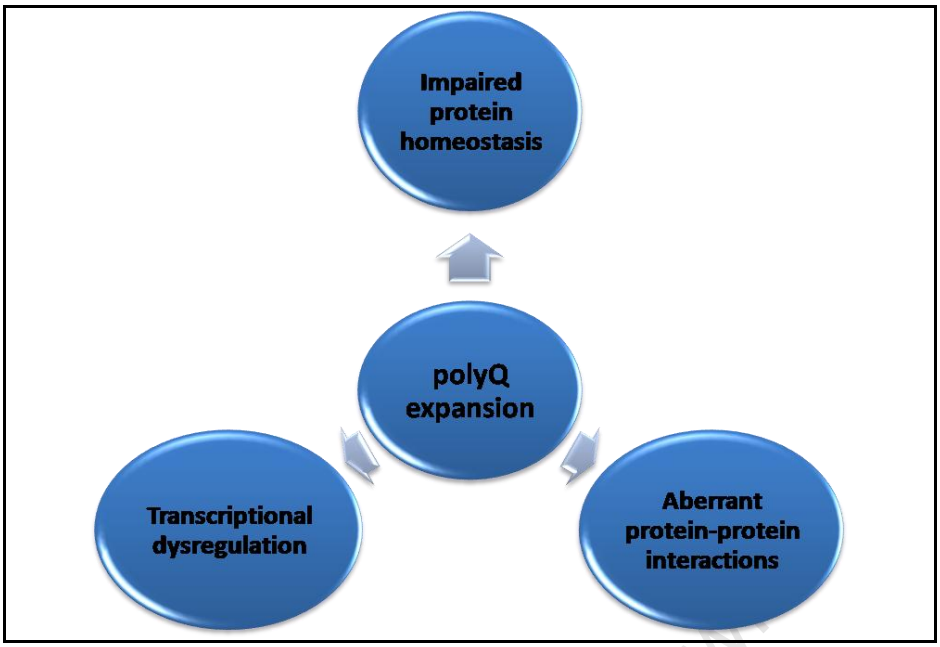


Figure 1.2 Possible mechanisms for SCA1 pathogenesis. The central aspect is the expansion of the polyglutamine tract. Protein misfolding as a result of polyglutamine expansion is believed to lead to impaired protein homeostasis, aggregation and thus neuronal cell death. On the other hand polyglutamine expansion has been shown to alter protein-protein interactions that may be necessary for cellular homeostasis and finally, ataxin-1 in its mutant form may lead to transcriptional dysregulation.

One of the important proteins identified so far is Brother of Ataxin-1 (Boat), a related protein that shares the AXH domain of ataxin-1 (Mizutani *et al.* 2005). Using a *Drosophila* model, Mizutani and colleagues discovered mutant ataxin-1's association with Boat modulates its cytotoxicity. In a SCA1 mouse model, Boat expression was found to be significantly reduced in Purkinje cells prior to the appearance of NIIs and dendritic thinning. The reason for this reduced expression is unknown but as an early event in SCA1 pathogenesis it could result in increased amounts of self-associated mutant ataxin-1, thus aggravating the mutant protein's toxicity and contributing to cell degeneration (Mizutani *et al.* 2005).

In 2006, Lam *et al.* reported the existence of large stable complexes containing both wild-type and expanded ataxin-1 and the transcriptional repressor Capicua (CIC) (Lam *et al.* 2006). Using yeast-two-hybrid assays, it was determined that the interaction occurs via the AXH domain of ATXN1 and a conserved region of CIC. Results also showed that mutation of the S776 residue of ataxin-1 reduces the amount of expanded protein incorporated into ATXN1-CIC complexes. Thus, S776 may play a critical role in regulating the formation of these complexes. In a *Drosophila* model expressing expanded human ATXN1, Lam *et al.* showed a restoration of mutant phenotypes to wild-type by co-expressing *cic*, indicating that Cic protein modulates ataxin-1 toxicity; and that mutant ataxin-1 protein interferes with the function of Cic as a repressor in *Drosophila* (Lam *et al.* 2006). From this data, it is clear that the wild-type form of ATXN1 interacts directly with CIC and modulates its activity, a function that is partially compromised by polyglutamine expansion. It is thus possible that expansion alters the conformational state of the protein, which in turn alters the conformational or functional state of the ATXN1-CIC complex (Lam *et al.* 2006).

In 2008, a two-pronged model of SCA1 neurodegeneration was proposed (Lim *et al.* 2008). Using a yeast-two hybrid screen Lim and colleagues identified two proteins whose interaction with ataxin-1 was dependent on polyglutamine expansion and phosphorylation of S776, both of which are critical for SCA1 pathogenesis. The two proteins, CIC and RNA-binding motif protein 17 (RBM17), form large distinct complexes with ATXN1. Both RBM17 and ATXN1 appear to be involved in the regulation of RNA metabolism, thus one potential function of the RBM17-ATXN1 complex may be RNA splicing (Lim *et al.* 2008).

As shown in figure 1.3, polyglutamine expansion in ATXN1 enhanced the formation of the RBM17-containing protein complex in both cell culture and SCA1 knock-in mouse cerebellum, providing a mechanistic explanation for the toxic gain-of-function.

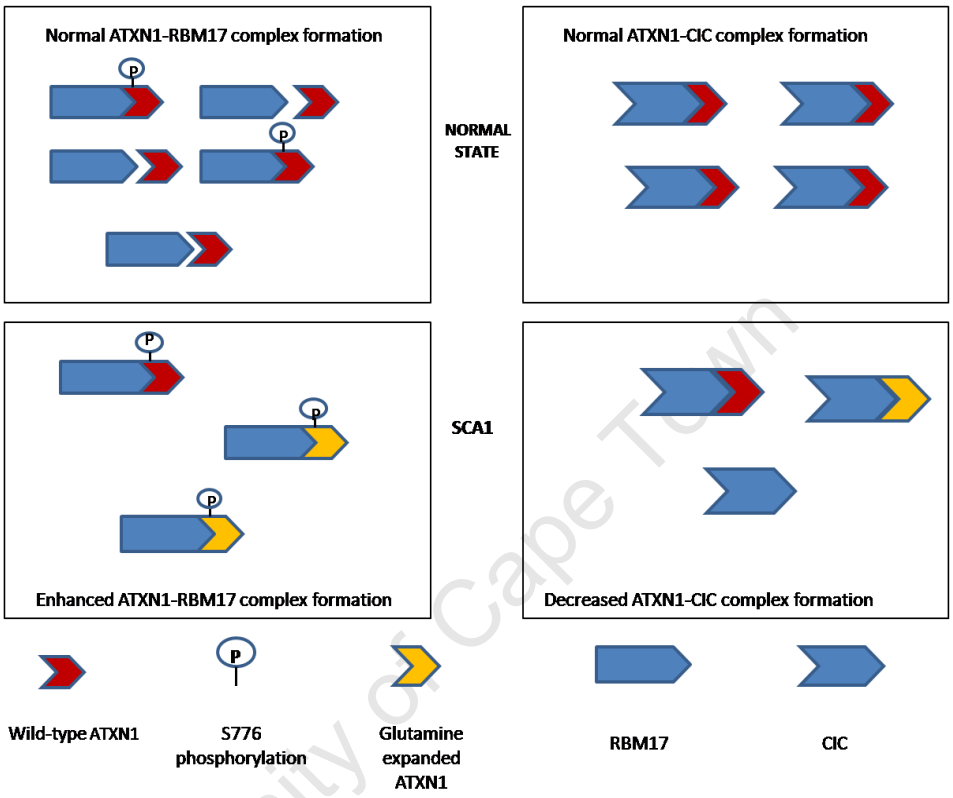


Figure 1.3 A two-pronged model for SCA1 pathology. In the normal state, there are at least two distinct and mutually exclusive ATXN1-associated complexes *in vivo*. The formation of ATXN1-RBM17 appears to require phosphorylation of ATXN1. Polyglutamine expansion in ATXN1 favours formation of the RBM17-containing complex, enhancing one endogenous function and contributing to neuropathology by a gain-of-function mechanism. Concomitantly, expansion decreases formation of the CIC-containing complex, resulting in a partial loss-of-function (adapted from Lim *et al*. 2008).

Mice lacking the wild-type protein but carrying the mutant form suffered more severe symptoms and greater lethality suggesting that wild-type ataxin-1 either protects against the mutant protein or is essential for neuronal integrity (Lim *et al.* 2008). The data also suggested that wild-type ataxin-1 competes with polyglutamine-expanded ataxin-1 to decrease formation of the toxic RBM17-containing complex (figure 1.3). Thus it is likely that the wild-type protein plays a protective role. Another possibility is that the ATXN1-CIC complex is neuroprotective. Lim and colleagues also showed that the wild-type protein is found in the CIC-containing complex while the expanded form preferentially forms a complex with RBM17.

The absence of two normal alleles results in an overall lower level of wild-type ataxin-1; thus less wild-type ATXN1-CIC complexes are formed, contributing to SCA1 neuropathology through a partial loss-of-function mechanism as shown in the figure (figure 1.3). The repressive activity of CIC was previously found to be greater in the presence of the wild-type ataxin-1 than with the expanded protein (Lam *et al.* 2006). The net effect of this reduced activity and the decreased co-repressive activity due to mutant ataxin-1 may also contribute to SCA1 pathogenesis (Lim *et al.* 2008). The authors suggest that this dual model may apply just as well to several other dominant neurodegenerative diseases caused by a gain-of-function mutational mechanism including other polyQ diseases, prion diseases and Alzheimer's disease.

Using a microarray system, a new study has investigated the difference between gain- and loss-of-function mechanisms by comparing the transcriptional defects in the cerebella of loss of Atxn1 function mice (*Atxn1*$^{-/-}$) to knock-in SCA1 mice (*Atxn1*$^{154Q/+}$) (Crespo-Barreto *et al.* 2010). The authors report a number of transcriptional changes common to both models, and speculate that these changes may indicate altered transcriptional functions of *ATXN1* in SCA1 pathogenesis. The study showed that most of the transcriptional changes go in the same direction in both *Atxn1*$^{-/-}$ and *Atxn1*$^{154Q/+}$ mice, a strong argument that part of the dysregulation in SCA1 might be explained by a partial loss-of-function of *ATXN1*. However, many of the transcriptional changes were found to be unique to the knock-in mouse model, which could be potentially related to the toxicity of polyglutamine-expanded ataxin-1 (Crespo-Barreto *et al.* 2010). Finally, the study demonstrated that over expression of ataxin-1-like, a related protein, can rescue some of the molecular and behavioural defects shown by the *Atxn1*$^{-/-}$ mice.

The authors suggest that ataxin-1 and ataxin-1-like share some normal functions, thus ataxin-1-like compensates for loss of ataxin-1 function in mice expressing the polyglutamine-expanded ataxin-1. *Atxn1L* was shown to be a functional homolg of *Atxn1* in Cic-mediated transcriptional repression (Crespo-Barreto *et al.* 2010). In addition to previous data (Lim *et al.* 2008), this study supports a two-pronged mechanism for SCA1 pathogenesis where enhanced toxic gain-of-function of mutant ataxin-1 leads to neurodegeneration, while a simultaneous loss-of-function of other endogenous protein complexes (ATXN1-CIC), contributes to the disease phenotype (Crespo-Barreto *et al.* 2010).

1.5 Introduction to RNAi technology

In 1998, Fire *et al.* found that long double-stranded RNA (dsRNA) can be used to induce gene silencing in *Caenorhabditis elegans* (*C. elegans*) (Fire *et al.* 1998). Less than ten years later, Andrew Fire and Craig Mello were awarded the Nobel Prize in Physiology or Medicine for their work on gene silencing. This discovery has been described as one of the major breakthroughs in molecular biology in the last decade and was termed RNA interference (RNAi). It is a naturally occurring process conserved in eukaryotes that involves the use of small non-coding RNAs to suppress the expression of genes in a sequence specific manner. Since its discovery, several different methods have been developed to manipulate RNAi for use as both a research tool and as a novel therapeutic strategy for the treatment of cancer, infection and neurodegenerative disease. By knocking down expression of genes in a sequence-specific manner, RNAi can be used to determine gene function as well as the pathogenic mechanisms of various diseases.

Small RNAs have been found endogenously expressed in several organisms including mammals, *Drosophila melanogaster* (*D. melanogaster*) and *C. Elegans* (Castanotto and Rossi 2009). Table 1.1 shows further details on these known small RNAs. One group is characterised by their interaction with the P-element induced wimpy testis (PIWI) clade of Argonaute proteins, and are thus known as PIWI-interacting RNAs (piRNAs). So far, these have only been identified in germline cells and originate from transposons, viruses and repetitive sequences. More recently another class of endogenous small interfering RNAs (endo-siRNAs or esiRNAs) has been identified in the gonads and somatic tissue of *D. melanogaster* and in mouse oocytes. They are proposed to regulate retrotransposon movement in mice.

Other small RNA families (Table 1.1) including repeat-associated siRNAs (ra-siRNAs), tiny non-coding RNAs (tncRNAs), *trans*-acting siRNAs (ta-siRNAs) and scan RNAs (scnRNAs) have been found in other organisms but as yet, not in mammals (Castanotto and Rossi 2009).

Table 1.1 Cellular small RNAs involved in gene silencing

Class	Size (nt)	Function	Mechanisms	Origin	Organisms found in
siRNAs	21-25	Regulating gene expression, providing antiviral response, restricting transposons	RNA degradation, transposon restriction	Intergenic regions, exons, introns	*C. elegans, D. melanogaster, Schizosaccharomyces pombe, Arabidopsis thaliana, Oryza sativa*
endo-siRNAs	21-25	Restricting transposons, regulating mRNAs and heterochromatin	RNA degradation	Transposable elements, pseudogenes	*D. melanogaster*, mammals
miRNAs	21-25	Regulating gene expression at the post-transcriptional level	Blocking translation, RNA degradation	Intergenic regions, introns	*C. elegans, D. melanogaster, S. pombe, A. thaliana, O. sativa,* mammals
piRNAs	24-31	Regulating germline development and integrity, silencing selfish DNA	unknown	Defective transposon sequences and other repeats	*C. elegans, D. melanogaster, Danio rerio,* mammals
ra-siRNAs	23-28	Remodelling chromatin, transcriptional gene silencing	unknown	Repeated sequence elements (subsets of piRNAs)	*C. elegans, D. melanogaster, S. pombe, Trypanasoma brucie, D. rerio, A. thaliana*
ta-siRNAs	21-22	*Trans*-acting cleavage of endogenous mRNAs	RNA degradation	Non-coding endogenous transcripts	*D. melanogaster, S. pombe, A. thaliana, O. sativa*
natRNAs	21-22	Regulating gene transcription at the post-transcriptional level	RNA degradation	Convergent partly overlapping transcripts	*A. thaliana*
scnRNAs	26-30	Regulating chromatin structure	DNA elimination	Meiotic micronuclei	*Tetrahymena thermophila, Paramecium tetraurelia*
tncRNAs	22	unknown	unknown	Non-coding regions	*C. elegans*

endo-siRNAs = endogenous siRNAs, miRNAs = microRNAs, natRNAs = natural antisense transcript RNAs, piRNAs = PIWI-interacting RNAs, ra-siRNAs = repeat-associated RNAs, scnRNAs = scan RNAs, siRNAs = short interfering RNAs, ta-RNAs = *trans*-acting RNAs, tncRNAs = tiny non-coding RNAs
(from Castanotti and Rossi 2009).

In humans, dsRNA is produced by a specific class of regulatory non-coding microRNAs (miRNAs). As shown in figure 1.4, synthesis of endogenous miRNAs begins with their transcription as primary microRNA (pri-miRNA) transcripts (Lee *et al.* 2002*a*). These are then processed in the nucleus by the Drosha/DGCR8 complex, into precursor miRNAs (pre-miRNAs) that are transported from the nucleus via an exportin-5-dependent mechanism (Lee *et al.* 2003, Yi *et al.* 2003). In the cytoplasm, pre-miRNAs are processed into short molecules by Dicer (an RNase III ribonuclease that generates RNA duplexes), approximately 22-nucleotides (nt) long, known as mature miRNAs. Endogenous miRNAs suppress gene expression by binding target messenger RNA (mRNA), blocking translation and facilitating degradation in processing bodies (P-bodies) (Sheth and Parker 2003, Hammond 2005).

One of the ways in which RNAi may be manipulated for gene silencing involves the use of artificial dsRNA molecules designed to specifically knock down the expression of mutant genes. This is most commonly mediated by siRNAs that are generated by the processing of complementary dsRNA templates into 21-23 nt duplexes (figure 1.4). The siRNAs consist of an 'antisense' (complementary to the target gene) and 'sense' (same sequence as the target gene) strand (Meister and Tuschl 2004, Hammond 2005). During assembly of the RISC, the siRNA is unwound in a strand specific manner and the antisense strand locates mRNA targets by Watson and Crick base-pairing. The antisense strand thus determines effective and specific silencing of the targeted mRNA. One of the components of RISC is Argonaute 2 (AGO2), an endonuclease which cleaves bound mRNA provided it is perfectly complementary to the guide strand (Liu *et al.* 2004, Hammond 2005).

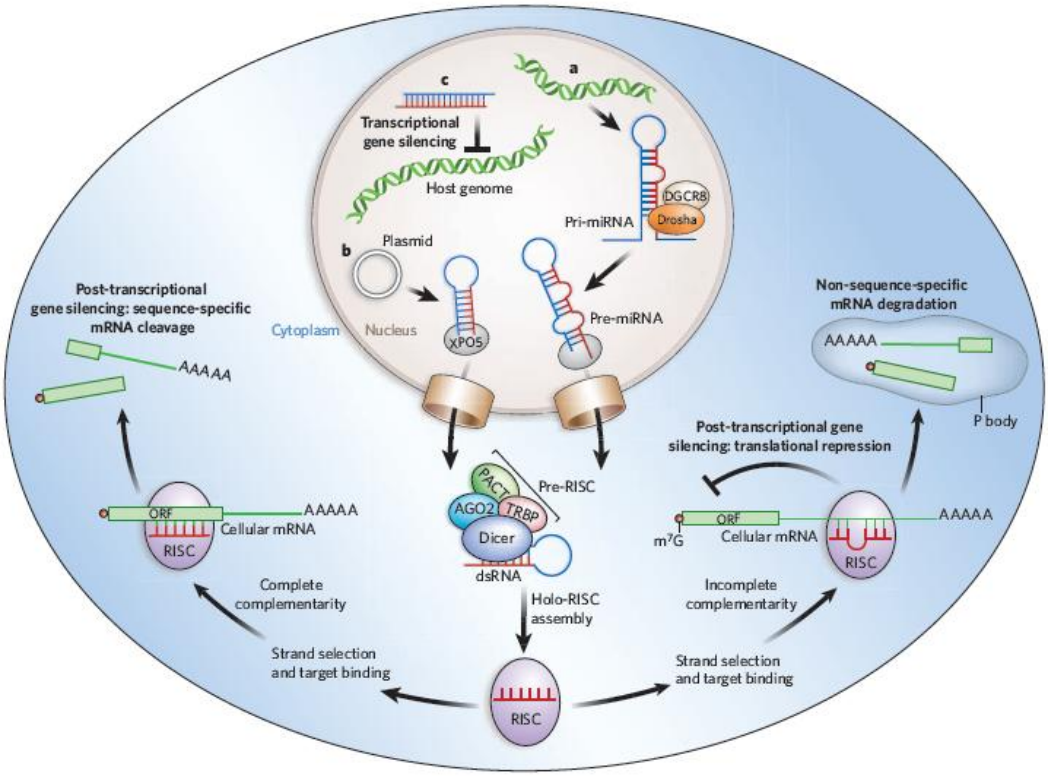


Figure 1.4 The mechanisms of cellular gene silencing. **a.** In plants and animals, pri-miRNAs are processed by Drosha and its partner DGCR8, into pre-miRNAs which are then transported to the cytoplasm by exportin 5 (XPO5). In the cytoplasm, pre-miRNAs are bound by a Dicer-containing pre-RISC and processed to yield the guide strand which is then loaded into holo-RISC (contains all the required components for gene silencing, including AGO2 the catalytic core). The guide strand binds to its complementary target sequence in the 3'UTR of cellular mRNA. If perfect complementarity is achieved (left of diagram), it triggers site-specific cleavage and degradation of the mRNA via the catalytic domain of AGO2. If base-pairing is incomplete (right of diagram) but includes pairing of the seed region (nucleotides 2-8 of the miRNA) with the target, translational inhibition occurs and may lead to non-sequence-specific degradation of the mRNA in P-bodies. **b.** Similarly to miRNAs, artificially transcribed shRNAs (in this case from a plasmid) are transported to the cytoplasm by XPO5 where the dsRNAs are recognised and processed by Dicer into 21-25 nt siRNAs that are loaded into RISC. The siRNAs are designed to target complementary sequences triggering their degradation by AGO2-mediated cleavage. **c.** In the nucleus, complementary siRNAs may trigger chromatin remodelling and histone modifications that result in transcriptional gene silencing. In mammalian cells, the mechanism involves the Argonaute-family proteins but the details are unclear. TRBP = HIV *tar*-RNA-binding protein, PACT = activator of protein kinase PKR, m^7G = 7-methylguanosine (adapted from Castanotto and Rossi 2009).

Several groups working on neurodegenerative disease to achieve targeting of transcripts using siRNAs have reported a number of challenges, the most important of which is their transient effect (Kim and Rossi 2007). Unlike in *C. elegans* where siRNAs are amplified by RNA polymerases, in mammalian cells which lack these enzymes, the effect of synthetic siRNAs is negated soon after the molecules are incorporated into the RISC complex (Cullen 2004). Other issues include side effects such as induction of interferon response, saturation of the endogenous pathway, delivery to certain systems or tissues and off-target effects. Different chemical modifications can be made to synthesised siRNAs to improve stability and efficacy and reduce off-target effects but they are still only effective in the short-term and require multiple applications (Kim and Rossi 2007, Castanotto and Rossi 2009).

Another major challenge with siRNAs is their delivery to the targeted cells and tissues since they cannot easily cross the cell membranes (Castanotto and Rossi 2009). Currently the main avenues are delivery of chemically modified siRNAs or the use of shRNA-encoding genes via recombinant viruses that generate siRNAs. Short hairpin RNAs (shRNAs) consist of sense and antisense strands separated by several nucleotides that form a non-complimentary loop. The shRNAs are delivered via plasmid or viral vectors into mammalian cells under the control of polymerase (Pol) II or III promoters (Kim and Rossi 2007). Pol III promoters are used where a high level of constitutive expression is required, while Pol II promoters can be used to drive tissue-specific expression. The products in both cases are efficiently incorporated into RISC. The use of promoter-expressed shRNAs or miRNAs can potentially mediate long-term silencing with a single application which addresses some of the issues mentioned above.

Targeting diseases of the CNS such as SCA1 further compounds the issue of safe and efficient delivery. The CNS is both structurally and functionally complex and is insulated from the vascular system by the blood-brain barrier (BBB) and the cerebral spinal fluid (CSF)-blood barrier. The use of recombinant viruses circumvents these problems since the viral vectors injected directly into the brain can yield expression of RNAi effectors driven by a tissue-specific promoter. Viral vectors can also be used for integrated permanent expression of the therapeutic gene, a requirement for chronic progressive diseases such as the dominant neurodegenerative disorders (Gonzalez-Alegre and Paulson 2007, Kim and Rossi 2007, Castanotto and Rossi 2009).

An important consideration in the choice of a vector is shRNA integration versus their episomal persistence within an expression cassette (Grimm and Kay 2009). Other considerations are the virus' inherent immunogenicity and the efficacy of the recombinant virus in a given cell or tissue. There are currently three major viral vectors for RNAi expression, adenovirus (Ad), lentivirus (LV) and adeno-associated virus (AAV). In spite of the success of viral delivery in mouse models of disease (Xia *et al.* 2004, Harper *et al.* 2005), major concerns exist pertaining to vector and RNAi safety. Viral vectors are potentially immunogenic and the risk of incurring mutations in viral sequences resulting in insertional mutagenesis or aberrant gene expression is relatively high (Castanotto and Rossi 2009). Other concerns involve the saturation of the RNAi machinery and the resultant toxicity, as well as the adverse effects of off-target silencing. These concerns may be addressed by adjusting the dosage or changing the promoter.

A study by Grimm and colleagues evaluated 49 AAV/shRNA vectors directed against 6 targets in livers of adult mice (Grimm *et al.* 2006). Thirty six of the 49 resulted in dose-dependent liver injury and 23 eventually resulted in death. The group proposes a competition model between processing of the over-expressed hairpins and endogenous miRNA to explain the observed toxicity. Exportin-5, a shared component of the miRNA and shRNA pathways, was identified as a limiting factor. The authors conclude that it is essential to monitor and control intracellular shRNA levels *in vivo* in order to achieve stable gene silencing while mitigating adverse effects (Grimm *et al.* 2006).

In 2008, another group developed shRNAs targeting conserved sequences in human HD and mouse HD homolog (HDh) mRNAs (McBride *et al.* 2008). The study screened 35 shRNAs *in vitro*, narrowing the candidate molecules to three for use in *in vivo* testing. In a knock-in HD mouse model, all three reduced HDh mRNA expression to similar levels. However, two of the three induced significant neurotoxicity in mouse striatum, possibly as a result of the observed higher levels of mature antisense RNA. The same sequences placed in an artificial miRNA expression system resulted in similar levels of reduced HDh mRNA expression with significantly reduced neurotoxicity. The authors conclude that miRNA-based systems may be more suitable for achieving RNAi in the brain (McBride *et al.* 2008).

A recent review summarises the significant improvements that have been made to original synthetic RNAi technologies in order to avoid the associated problems (Sibley *et al.* 2010). Strategies that involve the use of asymmetric interfering RNAs (aiRNAs), small internally segmented interfering RNAs (sisiRNAs) and many others, present significant advances in the understanding of RNAi and its exploitation as a therapeutic technique.

Further progress in this field is expected to improve the selection and optimisation of an effector, thus making therapeutic success more likely (Sibley *et al.* 2010).

1.6 RNAi and SCA1

Previously, proposed therapies for SCA1 and the other polyQs have targeted various downstream effects and aimed to enhance the solubility and turnover of the mutant protein by up-regulating chaperone proteins for example (Bonini and La Spada 2005). Prior to RNAi, targeted gene silencing involved the use of antisense oligonucleotide (AON) technology and catalytic nucleic acid biology using ribozymes (Wood *et al.* 2003). Ribozymes have been shown to be vulnerable to degradation by ribonucleases. AON technology has been useful in certain contexts but is plagued by low efficacy and a lack of specificity in mammalian systems. In spite of these disadvantages, AONs theoretically have the potential to target all individuals with a polyQ disorder (Scholefield and Wood 2010). Investigation into this technology is thus on-going.

RNAi as a therapeutic technique has shown tremendous promise in the search for a cure. Experiments on SCA1 mouse models have shown reversal of motor abnormalities and normalisation of Purkinje cell pathology. Xia and colleagues tested various AAV-expressed shRNAs targeting different sequences along the mutant transcript. Two of the constructs, flanking the CAG repeat, were found to mediate silencing in a cellular model and were then evaluated in a transgenic SCA1 mouse model. Results showed the elimination of NIIs in transduced cells correlating with improved neuropathology several days after inhibition of expression (Xia *et al.* 2004).

In another study, Zu and colleagues created a conditional SCA1 mouse model expressing a SCA1 human transgene (Zu *et al.* 2004). By suppressing the expression of transgenic mutant ataxin-1, SCA1 pathogenesis was shown to be completely reversible at early stages of disease and partially reversible even at a late disease stage. It is noteworthy however, that the effect of endogenous ataxin-1 was not examined.

Several studies have shown allele-specific RNAi to be effective in amelioration of disease phenotype in other polyQ disorders such as HD (Harper *et al.* 2005) and SCA3 (Miller *et al.* 2003, Li *et al.* 2004, Alves *et al.* 2008). Data thus supports the development of allele-specific silencing as a therapeutic strategy for SCA1. However, experiments are required to firstly determine the effect of this kind of treatment on a SCA1 knock-in mouse that more closely models the human condition, and secondly, on non-human primates to adjust the dosage required for the human cerebellum (Gonzalez-Alegre and Paulson 2007).

It is also important to note that although the function of wild-type ataxin-1 is unknown; its absence has been shown to cause a degree of neurological impairment in *Sca1* null mice (Matilla *et al.* 1998). These mice do not display any ataxia or major motor coordination abnormalities but they do show neurobehavioural abnormalities such as decreased exploratory behaviour and impaired performance on the rotor rod. It would be preferable to specifically target the mutant transcript with minimal knock-down of the wild-type using allele-specific RNAi. However, since mice lacking ataxin-1 show only subtle abnormalities it is possible that knockdown of both alleles in SCA1 patients may not result in major adverse effects as a result of loss of gene function (Matilla *et al.* 1998, Xia *et al.* 2004).

In 2008, Alves and colleagues reported the success of an allele-specific siRNA silencing strategy for SCA3 by targeting a single nucleotide polymorphism (SNP) in linkage disequilibrium with the disease-causing expansion in 70% of SCA3 patients (Alves *et al.* 2008). The approach developed used LV vectors encoding shRNAs to selectively target the mutant allele in a rat model of SCA3. More recently, this group has also investigated the global silencing of ataxin-3 with positive results (Alves *et al.* 2010).

1.7 South African perspective

A recent study has found that there may be success in applying an RNAi-based therapeutic approach for SCA7 patients in South Africa (Scholefield *et al.* 2009). All the SCA7 patients identified by the Human Genetics Division at the University of Cape Town (UCT) share the same haplotype, with a well-characterised SNP linked to the disease-causing expansion (Greenberg *et al.* 2006). With over 50% of patients heterozygous for this SNP, it provides a single nucleotide difference with which to discriminate between the wild-type and mutant transcripts. The study investigated allele-specific RNAi by targeting short sequences of *ATXN7* tagged to luciferase reporters, or full-length *ATXN7* cDNAs tagged to fluorescent reporter genes. An allele-specific effector was identified that has little effect on the wild-type protein and achieves strong silencing of the mutant. This suggests that targeting a single base pair difference, heterozygous in a significant number of patients, specifically where a founder effect exists, may be useful (Scholefield *et al.* 2009).

In 1997, a study was performed on 6 large SCA1 kindreds using three microsatellite markers D6S260, D6S89 and D6S274 (Ramesar *et al.* 1997). The authors reported the existence of two distinct founder effects in the Mixed Ancestry population in the Western Cape (Ramesar *et al.* 1997).

This study thus proposes a similar approach as that used for SCA7 (Scholefield *et al.* 2009); that is to develop a novel RNAi therapy for SCA1 in South Africa by targeting a single nucleotide difference linked to the disease-causing expansion.

1.8 Aims and Objectives

1.8.1 Aims

Allele-specific silencing has been used to target a linked SNP in the South African SCA7 patient population where a founder effect exists. It is proposed that the SCA1 patient population may also benefit from such an approach. Although the function of the ataxin-1 protein is not clearly defined, it is thought to play an essential role in cellular function, thus a method of preserving the wild-type protein would be desirable. As previously mentioned, two founder effects have been shown to exist in a specific population of South Africa. With advances in molecular genetics, the discovery of SNPs makes it possible to confirm the reported founder effects using SNPs in and around the *ATXN1* gene, instead of microsatellite markers. A SNP-based haplotype is more informative because of the increased availability of markers which makes it possible to narrow the region of analysis. In addition, genotyping SNPs may identify single nucleotide differences between the mutant transcript of ataxin-1 and its normal counterpart that may be potential targets for allele-specific RNAi.

Thus the aims of this project are firstly, to construct a more informative SNP-based haplotype to confirm the reported founder effects and secondly, to identify appropriate single nucleotide differences between the wild-type and mutant alleles of *ATXN1* and exploit these differences using RNAi to silence the mutant transcript.

1.8.2 Objectives

In order to achieve the above stated aims, the following objectives have been outlined:

A. Use SNP genotyping to identify possible founder haplotypes for the disease-causing allele.

B. Identify a SNP with sufficient heterozygosity in the patient population.

C. Design and screen tiled shRNAs against the wild-type and mutant targets *in vitro* to identify the most effective discriminating position of the SNP.

2. Materials and Methods

SECTION I: Construction of a SNP-based haplotype around the SCA1 locus

The Human Genetics Division at the University of Cape Town (UCT) has since the 1980s been involved in the collection of DNA samples from individuals clinically diagnosed with neurodegeneration. Informed consent (appendix C) was obtained and blood samples taken. DNA was isolated according to the protocols of the PUREGENE DNA Isolation Kit (*Gentra Systems*) outlined in "Rapid DNA isolation from 300 µl whole blood" or "DNA isolation from buffy coat prepared from 300 µl whole blood". Each DNA sample was assigned a laboratory reference code to ensure confidentiality, and available patient and family information was recorded onto the Microsoft Access Human Genetics Division Laboratory database.

The DNA samples were stored at -20°C for short term storage or at -80°C for long term storage.

As mentioned in Chapter 1, a previous study performed on six large SCA1 kindreds reported two distinct founder effects in the Mixed Ancestry families from the Western Cape (Ramesar *et al.* 1997). In this section, individuals from these kindreds were genotyped for several SNPs in order to construct a more informative SNP-based haplotype.

2.1 Selection of families for SNP haplotyping

In order to establish which haplotypes are associated with the disease-causing expansion in the Mixed Ancestry SCA1 population, a cohort of individuals was selected. The main criterion used was the presence of a complete triad, two parents (one of whom was affected) and an affected child. Eight triads were identified in five of the six Mixed Ancestry SCA1 families from the previous study, and individuals genotyped for their CAG repeat lengths and the selected SNPs.

2.2 Molecular genetic test for SCA1

Due to the instability of the expanded CAG repeat, different cells/tissues in an affected individual may carry alleles with slightly different repeat lengths. This makes the sizing of repeats somewhat subjective. Thus in order to confirm the affection status of individuals and the length of the CAG tract for purposes of this study, a PCR-based molecular genetic test was performed. DNA aliquots were taken from the stored stock samples and diluted to approximately 100 ng/μl concentration.

2.2.1 Polymerase chain reaction (CAG repeat)

PCR was performed using published primers (Dorschner *et al.* 2002) (appendix A) obtained from the molecular diagnostic laboratory in the Human Genetics Division, UCT. The PCR was performed as a standard reaction (appendix C), with the following thermal cycling conditions: 1 cycle of initial DNA denaturation at 95°C for 3 minutes; 28 cycles of (denaturation at 95°C for 60s, annealing at 61°C for 60s, elongation at 72°C for 90s); 1 cycle of final elongation at 72°C for 5 minutes.

2.2.2 Agarose gel electrophoresis

To confirm amplification, PCR products were resolved by electrophoresis through a 2% w/v agarose gel stained with ethidium bromide (EtBr) (*Sigma-Aldrich*) (appendix B). Ethidium bromide intercalates with DNA and fluoresces under ultra-violet (UV) light enabling visualisation of the DNA.

Cautionary note: Ethidium bromide has mutagenic properties. Care must be taken when working with this compound to ensure no skin contact. Wear impermeable gloves, a laboratory coat and other laboratory safety equipment where necessary.

Five microlitres of each PCR product were combined with 3 µl of agarose gel loading buffer (appendix B), loaded on to the gel and electrophoresed in 1 x Tris Borate-EDTA (TBE) buffer (appendix B) at 120 volts (V) for approximately 30 minutes. A size standard, GeneRuler™ 100bp Plus DNA Ladder (*Fermentas*, appendix C) was electrophoresed alongside to enable size estimation of the PCR products. After electrophoresis, the DNA was visualised under UV light using the uvipro gel documentation system (*uvitec*) and photographed by uvipro software (*uvitec*).

Cautionary note: Ultra-violet light degrades DNA. It is important to minimise exposure particularly if the DNA is to be used in any downstream experiments. Care must also be taken to avoid exposure to UV light, especially of the eyes. Wear impermeable gloves, a laboratory coat and goggles.

2.2.3 Genotyping on the 3100 ABI Genetic Analyzer

Once amplification was confirmed, PCR products were resolved on the ABI 3100 Genetic Analyzer (*Applied Biosystems*) to enable size estimation and subsequent repeat length calculation.

PCR products were prepared in a 96-well microtitre plate as follows: 1 µl PCR product, 0.3 µl Gene Scan ROX500 size standard (*Applied Biosystems*) and 8.7 µl Hi-Di formamide (*Applied Biosystems*) were added to each well. The mixture was denatured for 3 minutes at 95°C on a Touchdown Thermal Cycler (*ThermoHybaid*) and immediately cooled on ice prior to capillary based electrophoresis on the ABI 3100 Genetic Analyzer (*Applied Biosystems*). Results were collected using ABI PRISM 3100 Data Collection Software (v1.1, *Applied Biosystems*). Analysis of the raw data was performed using GeneMapper Software (v3.0, *Applied Biosystems*). Normal allele sizes were recorded. The sizes of the expanded alleles were estimated from the largest peaks visualised using GeneMapper.

The formula below was used to calculate the number of CAG repeats in each allele:

$$[(\text{amplicon size} - 163)/3 * 1.0678] + 1.115$$

This formula is used to adjust for anomalous migration of amplicons during capillary electrophoresis (Dorschner *et al.* 2002).

2.2.4 DNA sequencing

Several PCR products were randomly selected and sequencing performed to confirm the repeat sizes obtained by genotyping above.

Twenty microlitres of PCR product were combined with 3 µl of loading buffer, loaded onto a 2% w/v agarose gel stained with EtBr and electrophoresed in 1 x TBE for 30 minutes at 100V. The DNA was visualised under UV light and the bands quickly excised using a clean scalpel. The gel blocks were put into 1.5 ml eppendorf tubes and DNA extracted according to the manufacturer's protocol (appendix C) using the QIAquick Gel Extraction Kit (QIAGEN).

Three microlitres of purified extract were combined with loading buffer and electrophoresed to confirm minimal loss of DNA during gel extraction. Cycle sequencing (appendix C) was performed on the extracted DNA in the 5' to 3' direction using the ABI PRISM BigDye Terminator v3.1 Cycle Sequencing Kit (*Applied Biosystems*). (The SCA1 R primer (appendix A) carries a fluorescent tag, 6-FAM, and thus could not be used for sequencing).

Products of the cycle sequencing reaction were then purified by ethanol precipitation (appendix C) prior to resolution on the ABI PRISM 3100 Genetic Analyzer (*Applied Biosystems*). Results were collected using ABI PRISM Data Collection Software (v1.1, *Applied Biosystems*) and analysed using Sequencing Analysis Software (v3.7, *Applied Biosystems*). Electropherograms were examined manually using BioEdit Sequence Alignment Editor (v7.0.5.2, Hall 1999). The CAG repeats were counted and compared to the results obtained by genotyping.

2.3 SNP selection

To construct a SNP-based haplotype around the CAG repeat region in *ATXN1*, SNPs were selected using the following online databases: Entrez SNP in the National Centre for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov) and the Ensembl Genome Browser database (http://www.ensembl.org). A chromosomal region of approximately 450 kb spanning the *ATXN1* gene (6p23) was investigated for the presence of SNPs. The region was kept relatively small in order to reduce the probability of recombination events. SNPs were chosen according to the following criteria: a heterozygosity value approaching 50% and a minimum minor allele frequency (MAF) value of less than 10%. Heterozygosity values lower than 25% greatly reduce the chance of a triad within the same family yielding informative, in-phase genotypes. The MAF is indicative of a SNP's diversity in the population. The selection criteria were applied to data from both European and Yoruba populations in an attempt to represent the Mixed Ancestry population.

Table 2.1 shows the SNPs selected according to these criteria. Figure 2.1 shows the positions of the selected SNPs along the *ATXN1* gene, as well as one of the microsatellite markers used in the previous study and the position of the CAG repeat relative to the SNPs. The alleles indicated in the table are reference alleles from the NCBI website, according to the reverse sequence. However, note that sequencing is performed in the 5' to 3' (forward) direction and thus sequencing results are the complement of the alleles indicated here.

Table 2.1 SNPs selected for construction of a SNP-based haplotype

SNP	RefSNP accession ID	Reference alleles	Chromosomal location*
1	rs471716	C/T	6:16763242
2	rs4716071	A/ G	6:16433223
3	rs2075974	A/ G	6:16327330
4	rs180017	C /T	6:16306202
5	rs1042427	C /T	6:16295555

*taken from contig NT_007592.15, Genome Build: 37.1 (NCBI)

Alleles are from the reverse genomic DNA sequence as on the NCBI website.

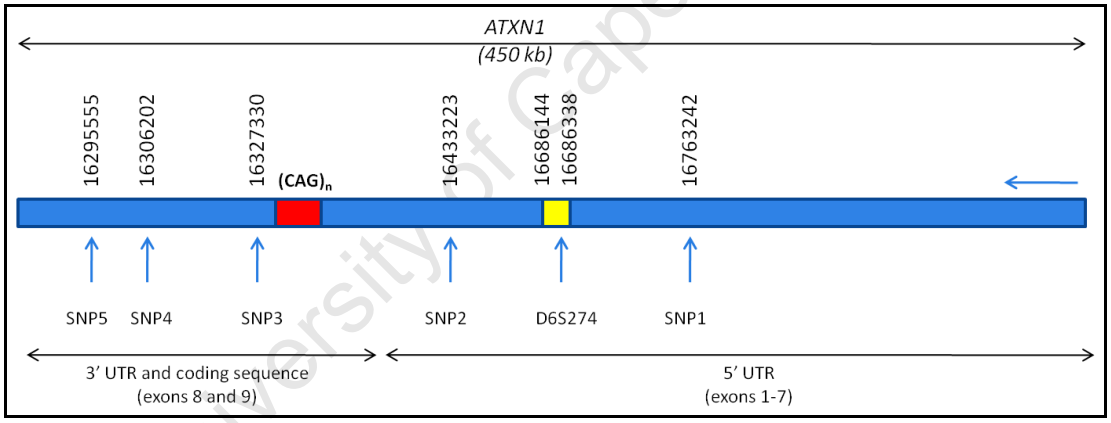


Figure 2.1 Positions of the selected SNPs along the *ATXN1* gene. Highlighted in RED is the position of the expanded CAG repeat, $(CAG)_n$. Highlighted in YELLOW is the position of the microsatellite marker D6S274, used in the previous study (Ramesar *et al.* 1997). The gene is transcribed in reverse, right to left of the figure as indicated by the blue arrow; the selected SNPs are thus numbered right to left (SNP1-rs471716, SN2-rs4716071, SNP3-rs2075974, SNP4-rs180017, SNP5-rs1042427).

2.4 SNP genotyping

In order to genotype the selected SNPs, it was necessary to design primers, perform PCRs and select a method for genotyping. Restriction enzyme digestion and single-strand conformational polymorphism analysis were used to determine SNP genotypes.

2.4.1 Primer design

Primers were designed to amplify the sequence of DNA spanning the SNP to be genotyped. The process of primer design was carried out as outlined:

- Primer3 (v.0.4.0, Rozen and Skaletsky 2000) was used to select possible primers from the sequence under investigation. The parameters were: length between 18 and 23 bp and melting temperature (T_m) between 50° and 58°C.
- The melting temperatures of each primer pair were calculated using OligoCalculator (Buehler, 1992) and matched to within 1°C of each other by manual adjustment of the primer length. The 3' end of each selected primer was either a "G" or a "C" nucleotide to enhance binding specificity.
- The potential of the primers to form secondary structures (hairpins, homodimers and heterodimers) was tested using OligoAnalyzer (v3.0, *Integrated DNA Technologies*).
- Finally primer sequences were subjected to a genome-wide sequence similarity search using the NCBI BLAST (Basic Local Alignment Search Tool: comparing nucleotide sequences, at http://blast.ncbi.nlm.nih.gov/Blast.cgi). This was to ensure that the primers were specific to the region of interest.

2.4.2 Polymerase chain reaction (SNPs)

PCRs were performed to amplify the target regions spanning each SNP. Optimisation for yield and specificity was carried out by altering one or more of the following parameters of the standard reaction (appendix C): annealing temperature (T_a), Magnesium Chloride ($MgCl_2$) concentration and thermal cycling conditions. For each PCR, the optimal T_a was determined with a temperature gradient experiment. The gradient started at least 5°C below and up to 5°C above the T_m of the primers as calculated by OligoCalculator. PCR products were electrophoresed and visualised as previously described. The optimal T_a was determined as the temperature at which the desired product yield was highest (brightest band) and showed the least amount of non-specific amplification (other bands). Where further optimisation was necessary, a $MgCl_2$ titration experiment was set up with the PCR containing 1 x Reaction Buffer without $MgCl_2$ (*Invitrogen*) and 50 mM $MgCl_2$ (*Invitrogen*) at final concentrations of 0.5 mM, 1 mM, 1.5 mM, 2mM and 2.5 mM in the total reaction volume. Thermal cycling conditions were also adjusted where necessary; longer annealing and elongation times and more cycles yielded more product.

Table 2.2 shows the primer sequences for each selected SNP, their optimal annealing temperatures and the expected amplicon size.

Table 2.2 Details of PCRs for amplification of SNPs

SNP	Primer sequences	T_a	Expected amplicon size (bp)
SNP1	**F** 5′ GTCCACTAGAGGCTGTAATG 3′ **R** 5′ CTAAGTGTCTGGAAACATGATC 3′	58°C	438
SNP2	**F** 5′ GAAATCTGCATGGAAGCCTG 3′ **R** 5′ CTGTTTCTAGATCTGATGCAC 3′	58°C	295
SNP3	**F** 5′ CTACAGCAGTCGTGATCCTTC 3′ **R** 5′ CTGTGTGTGGTCTGAATGAC 3′	58°C	210
SNP4	**F** 5′ ACGAGTGTCTGTCAGTGAG 3′ **R** 5′ TGAAGCTGGGTTCCTGATG 3′	56°C	361
SNP5	**F** 5′ TCAGCAGGAGGGCAACATTC 3′ **R** 5′ GTTATGAGTTCTGGGCAGCAG 3′	62°C	296

2.4.3 Restriction endonuclease digestion

SNPs 1, 3, 4 and 5 were genotyped by restriction endonuclease (RE) digestion. The restriction enzymes used were determined by analysis of each SNP amplicon using Webcutter 2.0 (Heiman, 1997). Table 2.3 shows the SNPs that were genotyped by RE digestion, the amplicon sizes and the sizes of the expected RE digestion products.

Table 2.3 SNPs genotyped by RE digest, showing amplicon and product sizes

SNP (genomic alleles)	Amplicon size (bp)	Expected RE digestion products (bp)	
SNP1 (C/T)	438	C allele: 370 bp and 68 bp*	T allele: 212 bp, 158 bp and 68 bp*
SNP3 (A/G)	210	A allele: 210 bp	G allele: 107 bp and 103 bp
SNP4 (C/T)	361	C allele: 361 bp	T allele: 185 bp and 176 bp
SNP5 (C/T)	296	C allele: 171 bp, 94 bp and 31 bp*	T allele: 232 bp and 64 bp*

*Fragment too small to visualise on an agarose gel

Note that the indicated alleles are from the reverse genomic DNA sequence

Table 2.4 gives the details of the RE digestions performed for each SNP. The amount of PCR product, the restriction enzyme used and the incubation temperature are shown. Each restriction digestion contained 5 units of the respective enzyme in a total volume of 20 µl, and was incubated overnight on a heating block.

Table 2.4 Set-up of RE digestions for SNPs 1, 3, 4 and 5

SNP	PCR product (µl)	Restriction enzyme	Temperature (°C)
*SNP1	7	*Alu*I	37
SNP3	10	*Stu*I	37
*SNP4	10	*Ava*II (*Eco*47I)	37
*SNP5	7	*Bsl*I	55

*After 1 hour of incubation, another 5 units of enzyme were added to the RE digestion and incubated overnight

Subsequent to RE digestion, 10 µl from each digestion were combined with 3 µl of agarose gel loading buffer, and electrophoresed through 2% w/v agarose gels as previously described. A size standard was electrophoresed alongside the digest products to enable size estimation. The gels were visualised under UV light and photographed as previously described.

SNP2 did not show a suitable variation at any restriction sites and was therefore genotyped using single-strand conformational polymorphism (SSCP) analysis.

2.4.4 Single-strand conformational polymorphism (SSCP) analysis

Genotypes for SNP2 were determined using SSCP analysis.

PCR was performed as a standard reaction (appendix C) with thermal cycling conditions adjusted as follows: 1 cycle of initial DNA denaturation at 95°C for 3 min;

25 cycles of (denaturation at 95°C for 15s, annealing at 58°C for 15s, elongation at 72°C for 45s); 1 cycle of final elongation at 72°C for 5 min. The shorter denaturation and annealing times served to reduce the amount of product from the reaction, giving clearer patterns for SSCP analysis.

Two 22 cm by 12 cm glass plates were cleaned thoroughly, first with ethanol and then with acetone. Plate glue was applied to the back plate to ensure that the gel matrix would stick to the plate (the front plate had blocks of Dyma tape used for well formation). A 0.5 mm spacer was sandwiched between the two plates. A 12% non-denaturing polyacrylamide gel solution was made, consisting of 5.3 ml of 40% w/v Acrylamide-piperazine diacrylamide (PDA) solution (appendix B), 8.5 ml of 0.75M Tris-formate buffer (appendix B) and 3 ml of 41% v/v glycerol.

Cautionary note: Acrylamide is a potent neurotoxin and irritant. Care must be taken to avoid any skin contact with, or inhalation of this compound. Wear the necessary laboratory safety equipment – laboratory coat, goggles, a mask and two pairs of impermeable gloves when working with this compound.

Polymerisation of the gel solution was achieved by adding 200 µl of 10% w/v ammonium persulphate (APS) (appendix B) and 20 µl of N'N'N'N'-tetramethylethylenediamine (TEMED) (*BDH Laboratory Supplies*). The total volume was poured between the glass plates and allowed to polymerise for at least 2 hours before use.

After separation of the plates, the back plate (gel plate) was placed on the LKB 2117 Multiphor II Electrophoresis Unit (*Pharmacia Biotech*). For each gel, 6 filter strips (approximately 25 cm x 5 cm) were soaked in 1 x Tris Borate (TB) buffer (appendix B) and 3 strips overlaid on each side of the gel. The PCR products were diluted 10-fold. One microlitre of the diluted PCR product was combined with 3 µl of SSCP loading buffer (appendix B) and loaded into the wells in the gel. Electrophoresis was performed at 5°C for 2 hours at 355 V using the Multiphor unit (*Pharmacia Biotech*).

The gels were then rinsed in distilled water to remove excess buffer. DNA was visualised by silver staining carried out as outlined. The gels were firstly incubated in a 0.1% w/v silver nitrate solution (appendix B) for 15 minutes, and then rinsed thoroughly with distilled water to remove any excess solution. Thereafter, they were incubated in a solution containing 0.375M sodium hydroxide and 0.8% v/v formalydehyde (appendix B) for 5-10 minutes until the staining was sufficiently developed. The gels were photographed using a digital camera.

2.4.5 DNA sequencing for SNP2

DNA sequencing was required to distinguish between the patterns shown by SSCP analysis and determine their particular genotypes. Sequencing was performed as previously described with two modifications. Firstly, the PCR products were sequenced directly without gel extraction as they are specific to the region of interest; and secondly in both the forward and reverse direction using the F and R primers.

2.4.6 Family studies

In order to complete the construction of the haplotypes in the 5 families, several additional family members were genotyped. This was necessary to establish phase for some loci, due to uninformativity and missing genotypes. A total of 41 individuals from the 5 families, comprising 27 affected subjects, were genotyped for the construction of the SNP-based haplotype. Family pedigrees were drawn using Cyrillic software (v2.1). Segregation analysis was used to infer genotypes where necessary; and to construct haplotypes for each individual.

2.5 Statistical analysis

Statistical analysis was performed using anonymous control DNA samples from individuals of Mixed Ancestry origin available in the Human Genetics Division (UCT). Controls are ideally asymptomatic and are sourced from the general population. DNA samples from these controls were genotyped and used to establish whether Hardy-Weinberg equilibrium (HWE) exists for the selected SNPs; and to exclude the possibility of the disease-associated haplotype occurring at a high frequency in the background Mixed Ancestry population.

2.5.1 Application of Hardy-Weinberg equilibrium

The phenomenon of HWE was first described in 1908. In the absence of migration, mutation, natural selection and assortative mating, genotype frequencies at any given locus are a simple function of allele frequencies (Wigginton *et al.* 2005). A deviation from HWE at a given marker suggests a change in population structure such as inbreeding, population stratification or it may indicate genotyping problems. In a sample of affected individuals, deviation from HWE suggests an association between the particular marker and disease susceptibility (Wigginton *et al.* 2005).

It is a well-established practice to check whether observed genotypes conform to Hardy-Weinberg expectations. For a locus with two alleles: A and C, with frequencies p and q respectively, where p + q = 1, HWE exists when:

$$p^2 + 2pq + q^2 = 1.$$

The expected frequencies of the genotypes AA, AC and CC are p^2, 2pq and q^2 respectively. A chi-squared test is normally used to indicate whether the data is significantly in disequilibrium. The null hypothesis is that the frequencies are in HWE, and is approximated with a continuous distribution. For a small sample size and discrete data such as genetic markers (Rohlfs and Weir, 2008), an exact test rather than a chi-squared test is required to make a more accurate analysis possible. An online webpage, http://ihg2.helmholtz-muenchen.de/cgi-bin/hw/hwa1.pl was used to perform the Hardy-Weinberg exact test to determine whether SNP genotype data fits HWE in the Mixed Ancestry population group.

2.5.2 PHASE analysis

The founder haplotypes identified in the Mixed Ancestry population (Ramesar *et al.* 1997) were investigated in the background population. For phase-unknown data from the unrelated controls, it was necessary for haplotypes to be inferred. This may be difficult due to the number of possible haplotypes which increases as a power of 2 with each additional SNP included in the analysis (Adkins, 2004). Several programs are available for this kind of analysis and have been shown to be highly accurate. PHASE was selected for this study.

PHASE (v2.1) is freely available at http://www.stat.washington.edu/stephens/software. The program uses a Bayesian statistical method to reconstruct haplotypes from population genotype data and is able to deal with SNP, microsatellite and other multi-allelic loci in any combination, as well as missing data. The Bayesian method treats unknown haplotypes as random quantities and combines previous information (what kind of patterns one would expect in a population sample), with the probability of the haplotype in the observed data (Stephens and Donelly, 2003). In this way it is possible to calculate the conditional distribution of the unobserved haplotypes (or haplotype frequencies) given the observed genotype data.

The format of the input file is explained in detail in the instructions in the program package. The input file specifies how many individuals are to be analysed, how many loci each individual has been genotyped at, the type of loci (SNP or microsatellite) and the genotype at each. PHASE uses an iterative scheme to perform inference. Parameters to control the number of iterations performed can be added to the input line, otherwise default values are used. The basic command line is as follows:

```
./phase <filename.inp> <filename.out>
```

Analysis was performed separately on the affected individuals and the controls respectively, using the following commands:

- **phase -s500 -f1 -kknown.txt patientonly.inp patientonly.out**
- **phase –s500 -f1 cononly.inp cononly.out**

The command line begins with "`phase`" which is used to invoke the program. The `-s` option is used to set the value of the seed of the pseudo-random number generator and must be a positive integer. To obtain reliable results, it is advisable to run the algorithm several times with different seeds and then check consistency across runs. The –f option specifies the format of the input file, `-f1` indicates that genotypes are listed on a single line for each individual. The `-k` option allows the user to specify known phases; "`known.txt`" is the file containing this information for the affected individuals. The command lines also specify the input files ("patientonly.inp" and "cononly.inp") and their respective output files. The program gives several output files, the first of which is a summary of the haplotype estimates for each individual. The other files give detailed results and estimates of the sample haplotype frequencies which is the required output in this case.

In this study, PHASE analysis was performed on the controls in order to determine the frequency of the identified disease-associated haplotypes. The haplotypes frequencies were then extrapolated to estimate the frequencies in the background population. Analysis was also performed on the patients in order to provide a more direct comparison of haplotype frequencies between the two groups, and to serve as a control by ensuring that the program selected the correct disease-associated haplotypes occurring as the most frequent or with a high frequency compared to the controls.

SECTION II: Allele-specific silencing

The use of allele-specific silencing for dominantly inherited disorders has shown some success. This approach may be particularly useful where a founder effect exists such that the disease-causing mutation is linked to a single SNP in a large number of affected individuals (Scholefield *et al.* 2009). In section I, a SNP-based haplotype was constructed and two SNPs were identified with greater than 50% heterozygosity in the SCA1 affected individuals. Short hairpin RNA effector molecules were designed against targets incorporating the selected SNP. Subsequently a dual luciferase reporter assay was used to screen the shRNAs in order to test the feasibility of this SNP as a target. The flow diagrams below outline the steps performed.

2.6 Plasmid construction/cloning (shRNA)

In order to construct a shRNA-expression system, several steps were performed as outlined in the flow-diagram (figure 2.2).

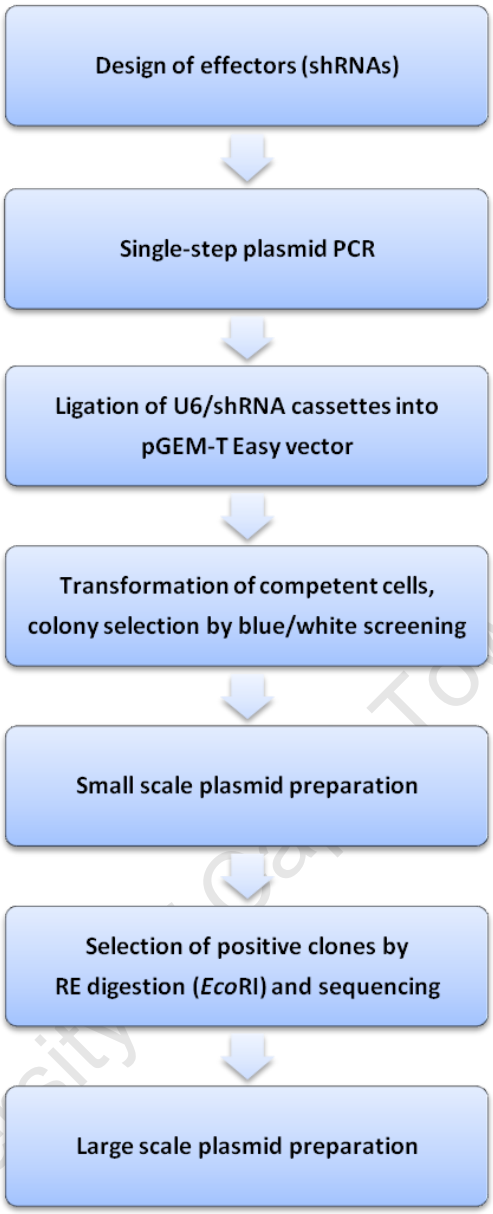


Figure 2.2 Flow-diagram outlining the construction of a shRNA-expression system.

2.6.1 Effector design

The shRNA effector molecules were designed based on the original expression vector system developed by Brummelkamp and colleagues (Brummelkamp *et al.* 2002). Their system used a Pol III promoter which produces small transcripts with a well-known start site and a five thymidine (T) termination signal. For this study, a U6 Pol III promoter from a pTZU6+1 vector (Lee *et al.* 2002*b*) was used to drive the production of small transcripts. The effectors were designed with 19 nucleotides (nt) complementary to the target sequence, defined as the antisense/guide strand; followed by another 19 nt of reverse complementary sequence, the sense/passenger strand (appendix A – oligos for shRNA construction). This sequence was interrupted by a 9 nt loop sequence (the miR-122 loop sequence). The antisense strand terminates with five Ts that result in a 2 nt 3' overhang because termination site cleavage occurs after the second U in the transcript (in the mRNA transcripts, genomic Ts are transcribed to Us and not As) (Baer *et al.* 1990). The favoured thermodynamics of the entire sequence predict the formation of a short hairpin loop when transcribed (see generalised structure in figure 2.3).

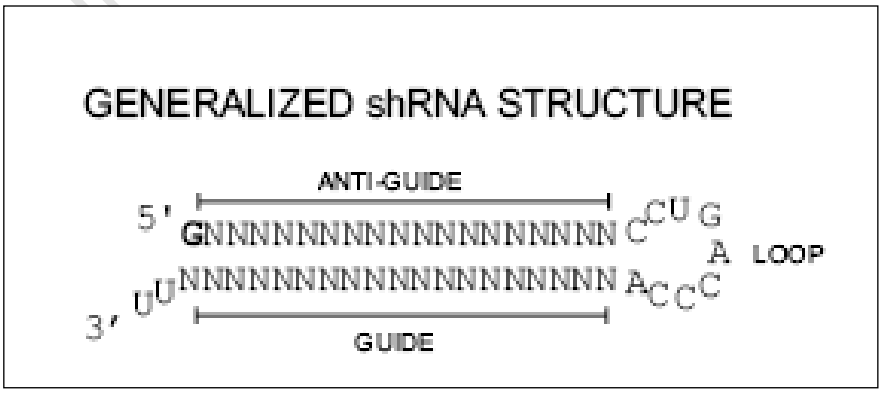


Figure 2.3 Diagram showing the generalised structure of predicted shRNA molecules

Cleavage of this dsRNA structure (figure 2.3) by Dicer is expected to yield the 21 to 22 nt RNA duplexes shown in figure 2.4 below. In order to bias antisense strand selection and incorporation into RISC, functional asymmetry was introduced by creating G:U wobbles with the addition of mismatches in the 3'end of the sense strand (Schwarz *et al.* 2003). Based on the predicted antisense strand, mismatches to the wild-type allele were introduced consecutively from position 9 to position 16; where position 1 is defined as the first nt of the 5' end of the antisense strand assuming a 21 nt duplex.

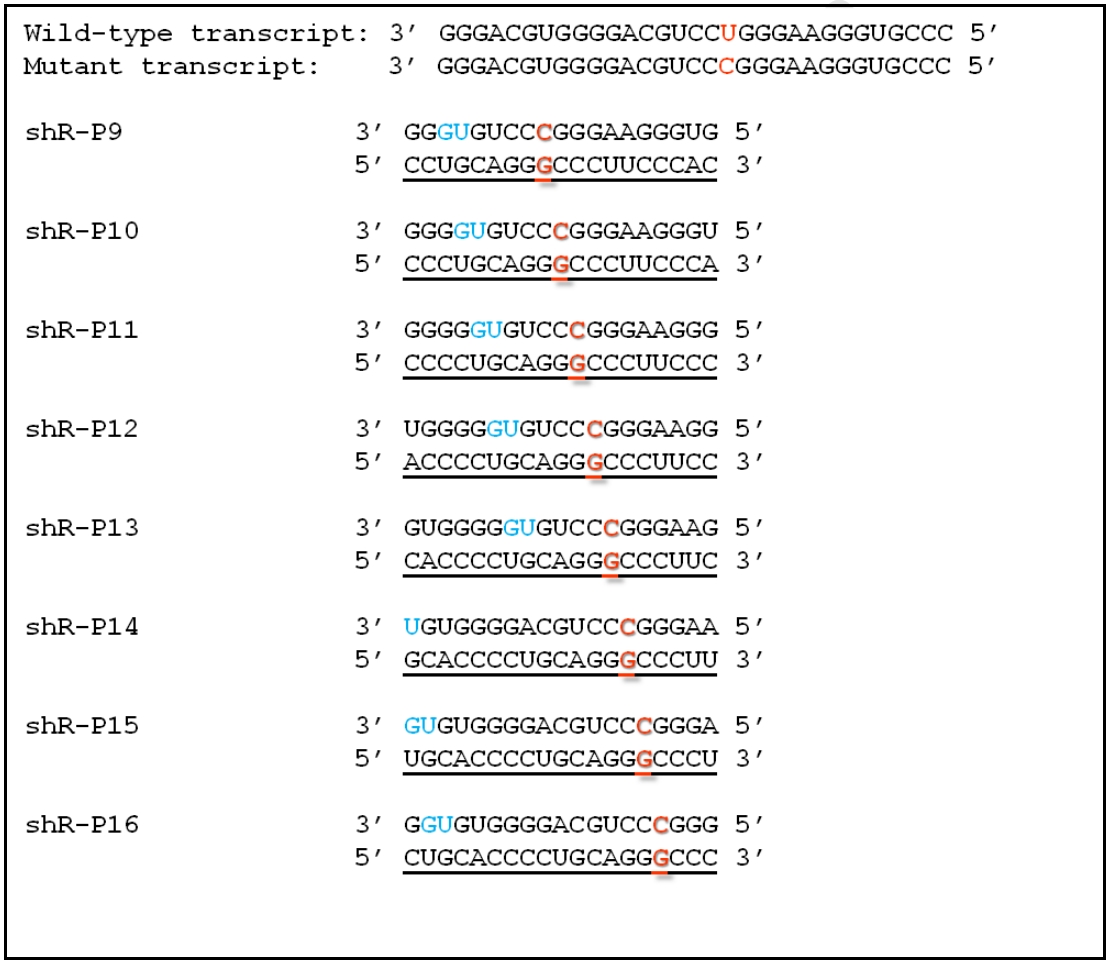


Figure 2.4 Panel of predicted siRNA molecules designed for allele-specific silencing. The target SNP is highlighted in **red**. Destabilising G:U wobbles are in **blue**. The underlined sequence is the predicted antisense (guide) strand designed to be perfectly complementary to the mRNA transcript.

2.6.2 Single-step plasmid PCR

Castanotto *et al.* describe a method for the construction of shRNA-expressing vectors in which the pTZU6+1 plasmid is used as a template to create U6/shRNA cassettes (Castanotto *et al.* 2002). This method was modified using a single-step PCR in order to construct U6/shRNA cassettes.

The single-step PCR was set up as follows: 50 pg of pTZU6+1 plasmid, 10 pmol U6+1 universal F primer (appendix A), 10 pmol long oligo R primer for each construct (appendix A), 0.2 mM dNTPs (*Bioline*), 1 x GoTaq Buffer (*Promega*), 2 units GoTaq (*Promega*); made up to a final volume of 25 µl with sH_2O. Thermal cycling conditions were as outlined (appendix C) with T_a of 60°C. To confirm amplification, 5 µl of the PCR products were combined with 3 µl of loading buffer, electrophoresed through a 2% w/v agarose gel and visualised as previously described.

2.6.3 Ligation

The PCR products (U6/shRNA cassettes) were ligated into the TA cloning vector, pGEM T-Easy (figure 2.5) using the pGEM T-Easy cloning kit (*Promega*). Ligation reactions were set up according to the manufacturer's instructions (appendix C) and left overnight at 4°C to ensure maximum transformants in the subsequent transformation step.

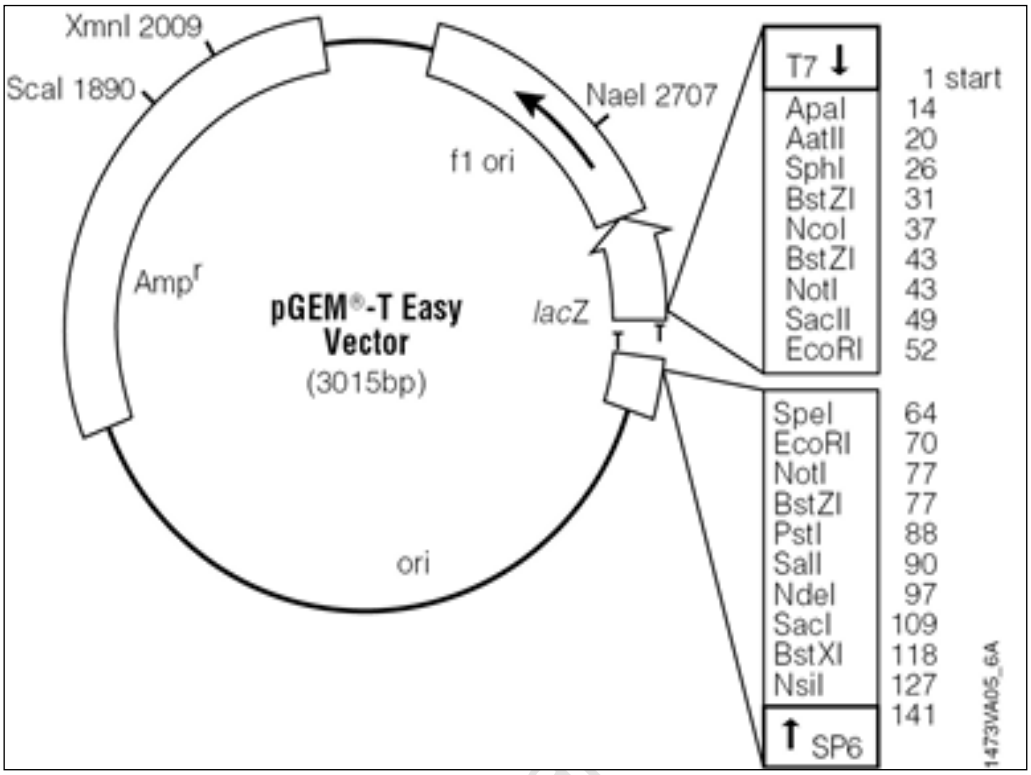


Figure 2.5 Diagram of pGEM®-T Easy Vector (*Promega*) showing the *lacZ* gene, the Ampicillin resistance (Amp^r) gene and the multiple cloning sites (www.promega.com).

2.6.4 Competent cells

Cloning is a technique that uses competent bacterial cells to replicate DNA of interest for use in subsequent experiments. Competency refers to the cells' ability to "take up" DNA under specific conditions, a process known as transformation. Thereafter, the transformed cells (transformants) are plated on suitable medium to encourage growth and colony formation. The resulting colonies serve as a source of the DNA of interest. All cloning experiments were performed using the *Escherichia coli* (*E. coli*) bacteria strain DH5α. Competent DH5α cells were prepared and stored at -80°C for use in the construction of effector and target plasmids.

2.6.4.1 Preparation of competent cells

DH5α cells were used to inoculate 5 ml of Luria broth (LB) (appendix B) in 15 ml Falcon tubes. The tubes were loosely capped to allow aeration and incubated overnight at 37°C with vigorous shaking. This starter culture was subsequently diluted 1/100 in 25 ml LB in 250 ml conical flasks and incubated for 1-2 hours at 37°C until log phase. The flasks were loosely covered with foil to allow aeration. Log phase was determined on a spectrophotometer by an O.D. reading of between 0.2 and 0.4 at 600 nm. The culture was transferred to 50 ml Falcon tubes and the cells collected by centrifugation (Beckman J2-21 centrifuge, *BeckmanCoulter*) at approximately 5000 rpm for 5 minutes at 4°C. **Note: Cells were kept on ice for all further steps.**

The cells were then re-suspended in half of the culture volume, ice-cold sterile 0.1M $CaCl_2$ and held on ice for 2 hours. The cells were centrifuged again and gently re-suspended in one tenth of the original culture volume, ice-cold sterile 0.1M $CaCl_2$. The cells were held on ice for a further 30 minutes then aliquoted (500 µl in 1.5 ml eppendorfs) and mixed with glycerol to a final concentration of 10% v/v. The aliquots of competent cells were labelled and stored at -80°C.

2.6.4.2 Transformation of competent cells

This method of preparation above is anticipated to produce cells with a transformation efficiency of approximately 1 x 10^7, suitable for most cloning experiments. It is important to calculate the transformation efficiency of competent cells in order to ensure successful cloning. This may be determined by transforming the competent cells with pUC19 plasmid DNA as follows: 100 µl of competent cells were combined with 1-10 ng of pUC19 plasmid DNA in 1.5 ml eppendorf tubes and held on ice for 30 minutes. A control tube with 100 µl of competent cells but no DNA was subjected to the same treatment.

The cells were heat-shocked at 42°C for 90 seconds and immediately transferred to ice for 2 minutes. Nine hundred microlitres of LB were then added and this mixture was incubated for 1 hour at 37°C in a shaker to allow expression of the antibiotic resistance gene. After incubation 100 µl each from the "expression mix" and the control were plated onto selective agar, and incubated at 37°C overnight.

Selective agar plates (appendix B) contain an antibiotic to facilitate the selection of positive transformants which form colonies due to the resistance conferred by the plasmid. Transformation efficiency was then calculated as follows:

number of colonies on plate/ng of DNA plated x 1000 ng/µg.

2.6.5 Transformation

Fifty microlitres of competent cells were combined with 5 µl of the ligation reaction products (U6/shRNA cassettes ligated to pGEM T-Easy) from section 2.6.3. Transformation was performed as previously described. Thereafter, 950 µl of LB were added to each tube and the tubes incubated.

2.6.6 Plating

The plasmid (pGEM®-T Easy, figure 2.4) carries the *lacZ* gene which encodes the α subunit of the β-galactosidase gene, while the host cells (DH5α) carry the gene encoding the γ subunit. Together the gene products form functional β-galactosidase enzyme that catalyses 5-bromo-4chloro-3-indolyl-β-Dgalactopyranoside (X-gal) (*Roche*) to an insoluble blue product. Ligation of DNA into the plasmid interrupts the *lacZ* gene such that no functional enzyme can be formed and therefore no blue product is produced.

All cells transformed with plasmid form colonies (ampicillin resistance conferred by plasmid). However, only those transformed with plasmid correctly ligated to the U6/shRNA cassette give white colonies. This serves to distinguish positive transformants, a technique known as blue/white screening.

Prior to plating, 40 μl of a 20 mg/ml solution of X-gal were spread on each plate to facilitate blue/white screening. Plates were allowed to dry in an incubator at 37°C. One hundred microlitres of the expression mix from each ligation were then spread on plates, which were incubated at 37°C overnight and then refrigerated for 4-5 hours to develop the blue colour.

2.6.7 Small scale plasmid preparation (Miniprep)

Small scale plasmid preparation, also known as miniprep, is a simple way to purify plasmid DNA (pDNA) from bacterial cells. This was important in order to purify the cloned plasmids from bacterial DNA.

From each plate, 4 or 5 white colonies were selected. Each colony (clone) inoculated a culture of 5 ml selective LB (100 μg/ml ampicillin) in 15 ml Falcon tubes. The tubes were loosely capped for aeration and incubated overnight at 37°C with vigorous shaking. After a maximum of 16 hours, the cultures were centrifuged at 4°C using a Beckman J2-21 centrifuge (*BeckmanCoulter*) to pellet the cells and the supernatant discarded. A labelled replica plate was made with some cells from each culture, incubated overnight at 37°C to allow growth, and refrigerated to serve as a source of the original colonies. Plasmid DNA was then purified from the pelleted cells using the QIAprep Spin Miniprep Kit (QIAGEN,) according to the manufacturer's instructions.

2.6.8 Quantitative and qualitative analysis of plasmid DNA

The pDNA obtained was quantified on a spectrophotometer and qualitatively analysed by agarose gel electrophoresis. A NanoDrop ND-1000 spectrophotometer (*NanoDrop Technologies*) was used to determine the concentration and purity of the pDNA. DNA concentration was measured in ng/µl and the protein/DNA ratio, which gives an indication of contaminating proteins and thus purity, was also recorded. Only samples with ratios between 1.8 and 2.0 were regarded as suitable for use in subsequent experiments.

Two microlitres of pDNA were then combined with 3 µl loading buffer, loaded on to a 0.8% w/v agarose gel and electrophoresed in 1 x TBE buffer at 80V for approximately one hour. After electrophoresis, the DNA was visualised under UV light as previously described. Plasmid DNA has various forms, circular, nicked/linear and super coiled. Each form is electrophoresed through an agarose gel at different speeds, and can thus be separated and visualised as indicated below in figure 2.6.

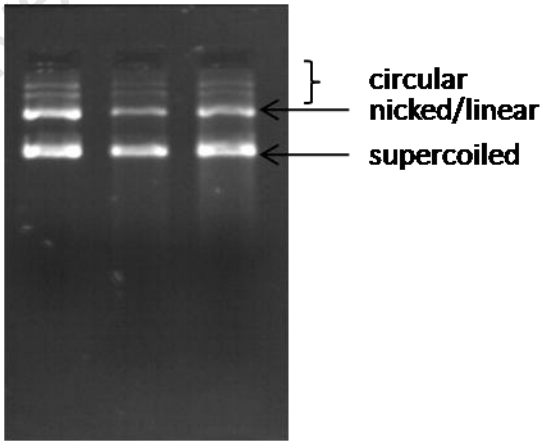


Figure 2.6 Gel photograph showing the different forms of plasmid DNA, circular, nicked/linear and super coiled. The DNA was electrophoresed on a 0.8% agarose gel stained with EtBr and visualised under UV light.

2.6.9 Restriction endonuclease digestion

To confirm the presence of the U6/shRNA cassette in a selected clone, a RE digestion was performed using *Eco*RI (*Promega*) on the purified pDNA. The reaction was set up as follows: 200 ng of pDNA, 12 units *Eco*RI, 1 x Buffer H (*Promega*) made up to a final volume of 20 µl with sH_2O. The reaction was incubated at 37°C for 2 hours. Ten microlitres of the digest product were then combined with 3 µl loading buffer, loaded onto a 0.8% w/v agarose gel and electrophoresed in 1 x TBE at 100V for 45 minutes. Two microlitres of the undigested pDNA were combined with 3 µl loading buffer and electrophoresed alongside the digested DNA for comparison. A size standard, GeneRuler™ 1kb DNA Ladder (*Fermentas*, appendix C) was included on the gel to enable estimation of the size of the insert. After electrophoresis, the DNA was visualised under UV light as previously described and photographed (figure 2.7).

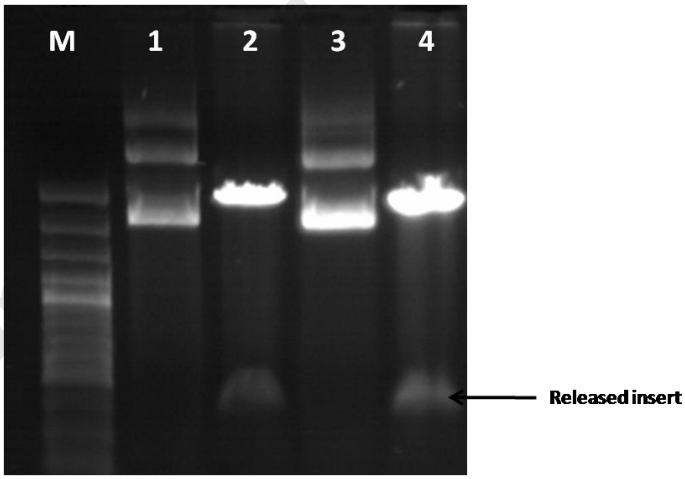


Figure 2.7 Gel photograph showing digested and undigested plasmid DNA (shRNA cassettes). M = molecular weight marker, lanes 1 and 3 = undigested plasmid DNA, lanes 2 and 4 = digested plasmid DNA with the released insert indicated. Plasmid DNA was electrophoresed on a 0.8% agarose gel stained with EtBr and visualised under UV light.

2.6.10 Sequencing of plasmid/insert

To confirm the sequence of the insert, DNA sequencing was performed as previously described. The cycle sequencing reaction (appendix C) was performed with approximately 500 ng pDNA from each clone and 24 pmol of the SP6 or the T7 primer (appendix A).

2.6.11 Large scale plasmid preparation (Midiprep)

Large scale plasmid preparation or midiprep, is a technique used to purify large quantities of pDNA from bacterial cultures. After a positive clone was selected and confirmed by RE digestion with *Eco*RI, and sequencing with SP6 or T7 primers, the corresponding colony was identified from the stored replica plate and 150 ml selective LB inoculated in a 250 ml conical glass flask. The flask was incubated overnight at 37°C with vigorous shaking and pDNA purified from the bacterial cells using the Midi/Maxiprep (QIAGEN,) according to the manufacturer's instructions.

2.7 Plasmid construction/cloning (targets)

Short target sequences of *ATXN1* incorporating the target SNP, were cloned into the psiCheck 2.2 vector (*Promega*). The flow-diagram below outlines the steps performed for the construction of the targets.

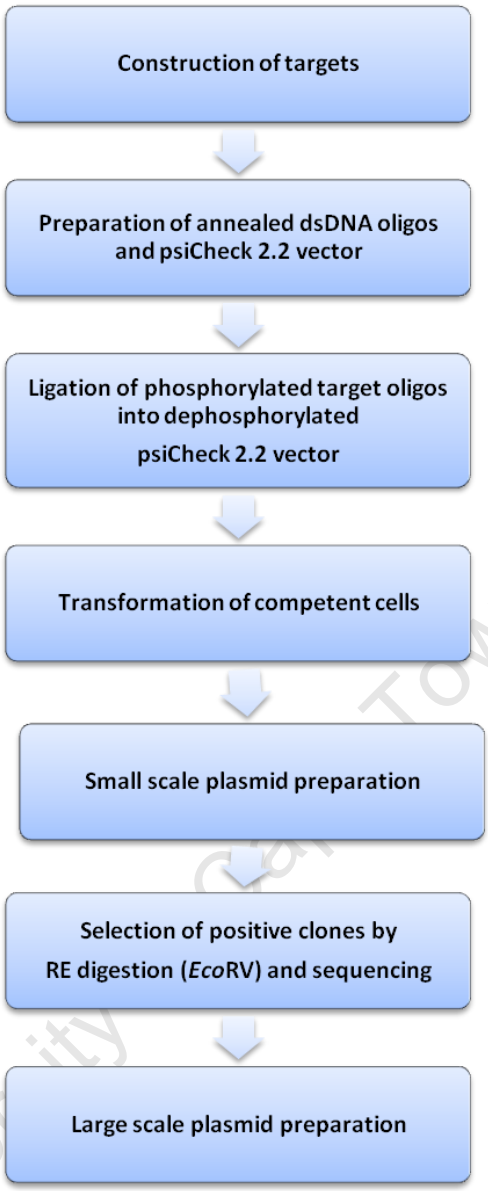


Figure 2.8 Flow-diagram outlining the construction of plasmid targets.

The psiCheck 2.2 vector expresses both the *Renilla* and the Firefly luciferase reporter genes (hRLuc and hFLuc respectively), shown in figure 2.9.

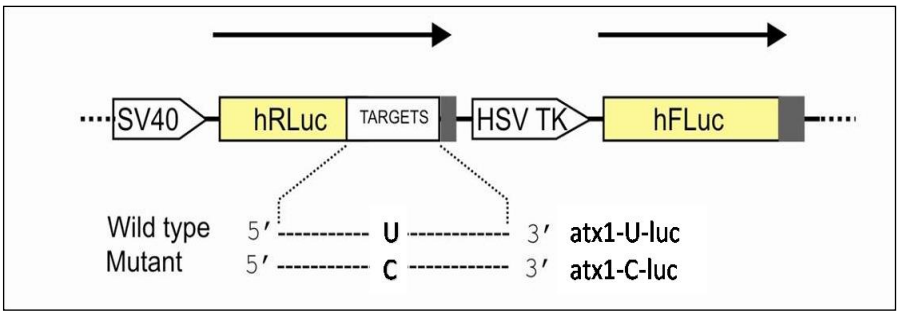


Figure 2.9 A schematic representation of the wild type (atx1-U-luc) and mutant (atx1-C-luc) dual luciferase reporter assay targets (adapted from Scholefield *et al*. 2009). The alleles indicated here are from the wild-type and mutant transcripts of *ATXN1*.
SV40 – Simian virus 40, hRLuc – *Renilla* luciferase gene, hFLuc – Firefly luciferase gene,
HSV TK – Herpes Simplex Thymidine Kinase

2.7.1 Preparation of annealed dsDNA oligos

The DNA oligonucleotides used for cloning the target plasmids were 62 nt long sequences incorporating the target SNP, and *Not*I and *Xho*I sticky ends (sequences in appendix A). The oligos were annealed and phosphorylated prior to ligation into psiCheck. Each oligo was phosphorylated in a separate reaction set up as follows: 100 μM oligo, 10 mM ATP (*Novagen*), 5 units T4 polynucleotide kinase (PNK) (*Novagen*), 1 x polynucleotide kinase buffer (*Novagen*), made up to 20 μl with sH_2O. The reactions were incubated at 37°C for 1 hour. After phosphorylation, complementary oligos were mixed to a final concentration of 5 μM each, incubated at 75°C for 10 minutes to inactivate the enzyme, and then annealed by slow cooling (i.e. the heating block was switched off and allowed to cool). Aliquots of the annealed dsDNA oligos were diluted 25-fold to a final concentration of 200 nM.

2.7.2 Preparation of dephosphorylated psiCheck 2.2 vector

In order to prepare the plasmid for ligation, a RE digestion was performed to give free ends, and dephosphorylated to minimise re-ligation.

The RE digestion was set up as follows: 2 μg of psiCheck 2.2 vector, 10 units *Not*I (*Promega*), 15 units *Xho*I (*Promega*), 1 x compatible buffer D (*Promega*) and made up to 30 μl with sH_2O. The reaction was incubated at 37°C for 2 hours. Thereafter, the following were added: 5 units Antarctic phosphatase (*New England Biolabs*), 1 x Antarctic phosphatase buffer (*New England Biolabs*) and made up to 40 μl with sH_2O. This mixture was incubated for a further 1 hour at 37°C, after which the phosphatase was inactivated by incubation at 65°C for 15 minutes. The total volume was then combined with 5 μl loading buffer, loaded onto a 0.8% agarose gel and electrophoresed in 1 x TBE at 80V for 1 hour. After electrophoresis, the high molecular weight band was visualised under UV light and quickly excised. DNA was purified from the gel block using the QIAQuick Gel Extraction kit (QIAGEN) according to the manufacturer's instructions (appendix C).

2.7.3 Ligation

The annealed dsDNA oligos were then ligated into the psiCheck 2.2 vector downstream of the *Renilla* reporter gene. A ligation reaction was set up between each of the phosphorylated dsDNA oligos and the linearised dephosphorylated vector, in a 1:3 ratio (vector:oligo). The reactions were set up as follows: 60 nM of psiCheck vector, 180 nM of dsDNA oligo, 5 units T4 DNA ligase (*Fermentas*), 1 x T4 DNA ligase buffer (*Fermentas*) and made up to 10 μl with sH_2O. A negative control reaction was also set up without ligase. The reactions were left overnight at 4°C.

2.7.4 Transformation

One hundred microlitres of competent cells were combined with 5 µl from each ligation reaction and transformation performed as previously described. Nine hundred microlitres of LB were added to each tube and incubated. Thereafter, 100 µl of the expression mix were plated on selective agar plates and incubated at 37°C overnight. After overnight incubation, the control plate should have none or very few colonies in comparison to the plates with transformants. This is an indication of successful ligation.

2.7.5 Small scale plasmid preparation (Miniprep)

Each colony on the transformed plate was picked to inoculate selective LB and small scale purification carried out on the bacterial cultures as previously described.

2.7.6 Quantitative and qualitative analysis of plasmid DNA

The purified pDNA was analysed quantitatively and qualitatively as previously described.

2.7.7 Restriction endonuclease digestion

To confirm the presence of the cloned target, a RE digestion was performed using *Eco*RV (*Roche*). The reaction was set up as follows: 200 ng of pDNA, 10 units *Eco*RV, 1 x buffer B, made up to a final volume of 20 µl with sH_2O. The reaction was incubated at 37°C for 2 hours. Digest products were then electrophoresed on a 0.8% w/v agarose gel, viewed under UV light as previously described and photographed (figure 2.10).

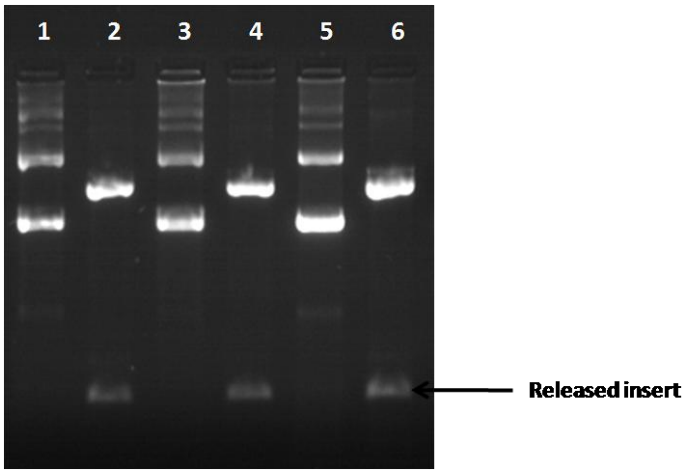


Figure 2.10 Gel photograph showing digested and undigested plasmid DNA (targets). Lanes 1, 3 and 5 = undigested plasmid DNA. Lanes 2, 4 and 6 = digested plasmid DNA with the released insert as indicated. Plasmid DNA was electrophoresed on a 0.8% agarose gel stained with EtBr and visualised under UV light.

2.7.8 Sequencing of plasmid/insert

Sequencing was performed directly on the pDNA to confirm the sequence of the insert as previously described. The cycle sequencing reaction (appendix C) was set up with approximately 500 ng of each of the purified pDNA and 24 pmol of psi-F or psi-R primer (appendix A).

2.7.9 Large scale plasmid preparation (Midiprep)

After identification of a positive clone by RE digest with *Eco*RV and confirmation by sequencing with psi-F or psi-R primers, the corresponding colony was identified from the stored replica plate and large scale preparation carried out on the bacterial cultures as previously described.

2.8 Cell-based assays

2.8.1 Maintenance of cell lines

All transfection experiments were performed in HEK293 cells. The HEK293 cell line is well-known and commonly used for *in vitro* experiments. Cells were cultured in Dulbeco's modified Eagle's medium (DMEM) (*Sigma-Aldrich*) supplemented with 10% v/v Fetal Calf serum (FCS) (*Sigma-Aldrich*), Penicillin-Streptomycin (*Sigma-Aldrich*) and Fungin (*Invitrogen*). The cells were maintained in a 37°C incubator with controlled carbon dioxide levels and medium replaced every three to four days.

2.8.2 Transfections

HEK293 cells were seeded in 24-well plates at approximately 120 000 cells per well. The plasmid transfections were performed using Lipofectamine2K (*Invitrogen*) according to manufacturer's instructions, with a total of 1 µg of DNA in a 5:1 ratio of effector: target in each well. A GFP-expressing plasmid (100 ng) was co-transfected with the target-expressing plasmid to give an indication of transfection efficiency. After 24 hours, the cells were transfected with the shRNA-expressing plasmid. A dual luciferase assay was performed 48 hours after the initial transfection.

2.8.3 Dual luciferase reporter assay

Reporter systems are commonly used to study gene expression and may be applied to intracellular signalling, mRNA processing and many others. The term "dual reporter" refers to the simultaneous expression and measurement of two individual reporter enzymes within a single system (www.promega.com).

One of these is the "experimental" reporter whose activity is correlated with the effect of the experimental conditions, while the other is the "control" reporter which provides a baseline response. Normalising the experimental reporter activity to the control reporter activity minimises experimental variability allowing more reliable interpretation of the results. The Dual-Luciferase® Reporter (DLR™) Assay System (*Promega*) measures the luciferase activities of firefly (*Photinus pyralis*) and *Renilla* (*Renilla reniformis*) from a single sample. The firefly luciferase reporter is measured first by adding Luciferase Assay Reagent II (LAR II). After quantifying the luminescence, this reaction is quenched and the *Renilla* reaction initiated by adding Stop and Glo® Reagent (www.promega.com).

Forty-eight hours after transfection, cells were washed in 1 x phosphate buffered saline (PBS) (*Sigma-Aldrich*) and harvested by incubating with 1 x passive lysis buffer (PLB) (*Promega*) for fifteen minutes. Forty micro litres of each sample were pipetted into individual wells of a white luminometer microplate. Forty microlitres of LAR II was added to each well, the plate was loaded into a Veritas Microplate Luminometer (*Turner Biosystems*) and firefly luciferase activity measured. Thereafter, the Stop and Glo® Reagent was added to each well and the plate re-loaded into the luminometer to measure the *Renilla* luciferase activity. Readings were recorded on a computer connected to the luminometer. *Renilla* luciferase/firefly luciferase ratios were analysed in relation to a non-specific control (i.e. shRNA molecule not complementary to the *ATXN1* gene).

3. Results

SECTION I: Construction of a SNP-based haplotype around the SCA1 locus

3.1 SNP genotyping results

3.1.1 Restriction endonuclease digestion

Following PCR amplification, four of the five selected SNPs were genotyped by RE digestion. Figure 3.1 shows examples of the results obtained for each SNP. All samples analysed for SNP1 showed the same genotype (figure 3.1 (a)). This SNP was deemed to be non-informative and was not included in further analyses. This result will be discussed further in a later section.

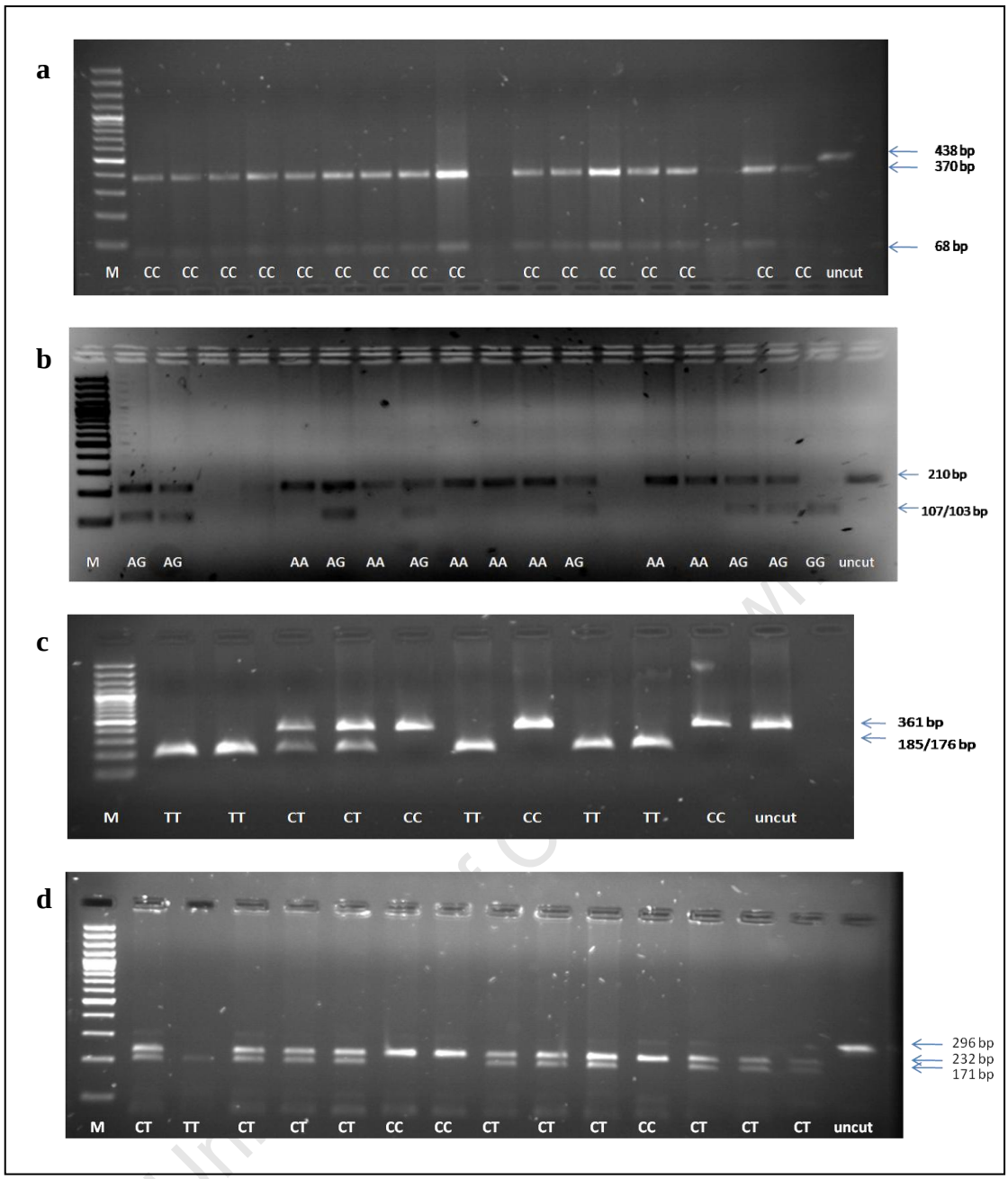


Figure 3.1 Gel photographs showing examples of SNP digest products.

(a) SNP1 PCR products after digestion with *Alu*I showing CC genotypes. (b) SNP3 PCR products after digestion with *Stu*I showing AA, GG and AG genotypes. (c) SNP4 PCR products after digestion with *Ava*II (*Eco*47I) showing CC, TT and CT genotypes. (d) SNP5 PCR products after digestion with *Bsl*I showing CC, TT and CT genotypes. M = molecular weight marker. All digest products were electrophoresed through 2% agarose gels stained with EtBr and visualised under UV light. Gel photo in (b) is inverted for a clearer view of digest products. Note that the indicated alleles are from the reverse genomic DNA sequence as on the NCBI website.

3.1.2 Single-strand conformational polymorphism

Due to the absence of a suitable restriction site, SNP2 was genotyped by SSCP analysis. Figure 3.2 shows an example of the results obtained from SSCP, with the genotype patterns determined after sequencing.

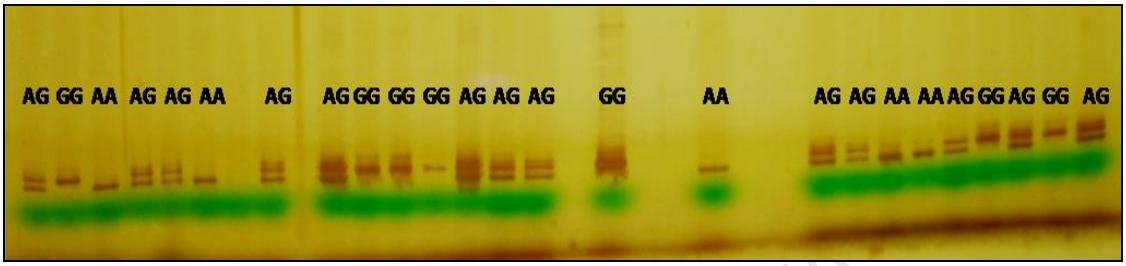


Figure 3.2 Gel photograph showing the SSCP banding patterns and respective genotypes for SNP2. The gel was stained with silver and photographed using a digital camera. Note that the indicated alleles are from the reverse genomic DNA sequence

3.1.3 SNP allele frequencies

Table 3.1 gives a breakdown and comparison of the allele frequencies for patients and controls.

Table 3.1 SNP allele frequencies, Patients and Controls

	SNP2		SNP3		SNP4†		SNP5*	
	T	C	T	C	G	A	G	A
Patients (N = 54)	0.41	0.59	0.9	0.1	0.69	0.31	0.55	0.45
Controls (N = 150)	0.37	0.63	0.67	0.33	0.39	0.61	0.67	0.33

Key:

= skewed distribution, N = total number of genotyped alleles for each SNP

†Controls N = 142; *Controls N = 132

Note that the indicated alleles are from the forward strand. Importantly, SNP3 and SNP4 highlighted in the table above (green box), show a significantly skewed distribution of the T allele in the affected individuals when compared to the control population ($p = 0.0003$ and $p = 0.0012$, respectively). A p-value less than 0.05 is considered significant.

3.2 Family studies

Eight triads were initially selected from five of the six kindreds of mixed ancestry investigated in the previous study (Ramesar *et al.* 1997): CA1, CA2, CA4, CA8 and CA11. It was necessary to genotype additional family members in order to establish phase thus expanding the triad. Cyrillic software was used to draw the family pedigrees. A total of 41 individuals were genotyped for CAG repeat length (normal and expanded alleles), and for the selected SNPs. Twenty-seven subjects carried an expanded repeat and the rest had two normal alleles. Normal repeat sizes ranged from 26 to 32 repeats while the expanded repeat sizes ranged from 47 to 66 repeats. All genotyping results were entered into Cyrillic and segregation analysis used to construct SNP haplotypes for each individual.

Pedigrees were drawn for the 5 families, figure 3.3 (a-e). Where a DNA sample was unavailable, genotypes were inferred if possible. The disease-associated haplotypes were highlighted as coloured bars. Note that the alleles indicated in the pedigrees are named from the forward strand.

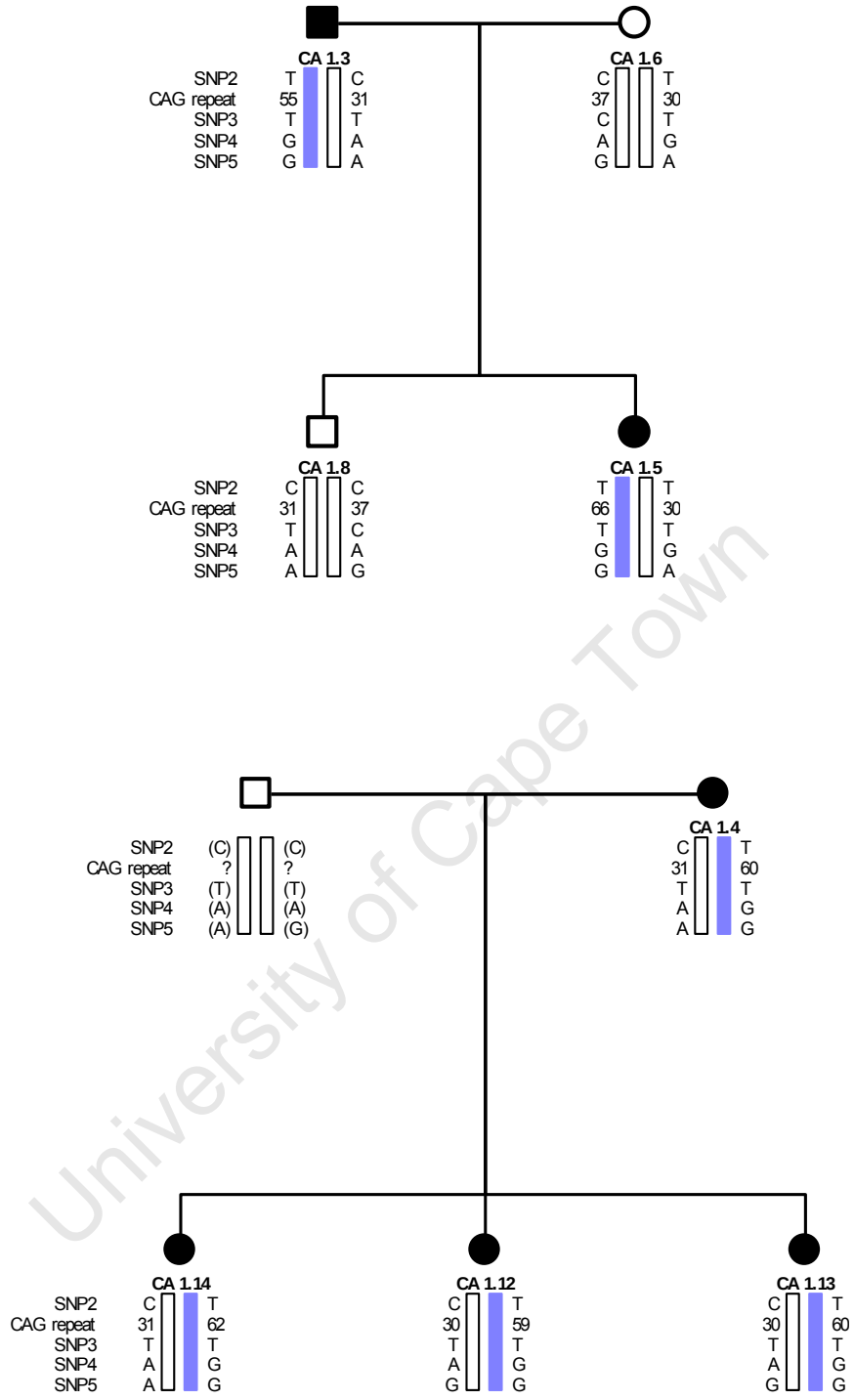


Figure 3.3 a) Pedigrees from Family 1.

Haplotypes constructed with SNP genotypes obtained from this study. Blue bars represent the haplotype associated with the CAG expansion. Circles represent females, squares represent males. Genotypes in brackets "()" are inferred. "?" represents unknown genotype.

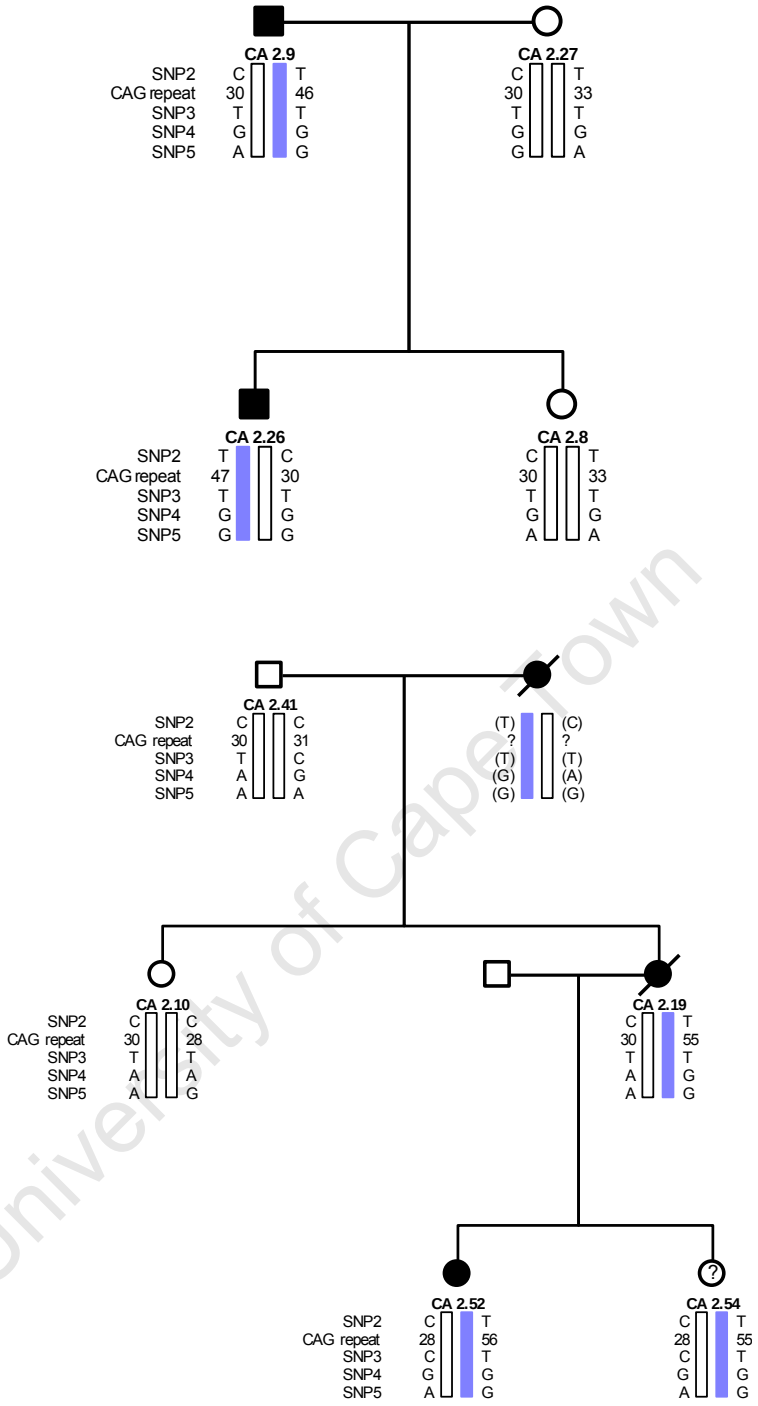


Figure 3.3 b) Pedigrees from Family 2.

Haplotypes constructed with SNP genotypes obtained from this study. Blue bars represent the haplotype associated with the CAG expansion. Circles represent females, squares represent males. Filled in symbols are affected individuals. Symbols with a "?" inside indicate unknown clinical status. A line through the symbol represents a deceased individual. Genotypes in brackets "()" are inferred. "?" represents unknown genotype.

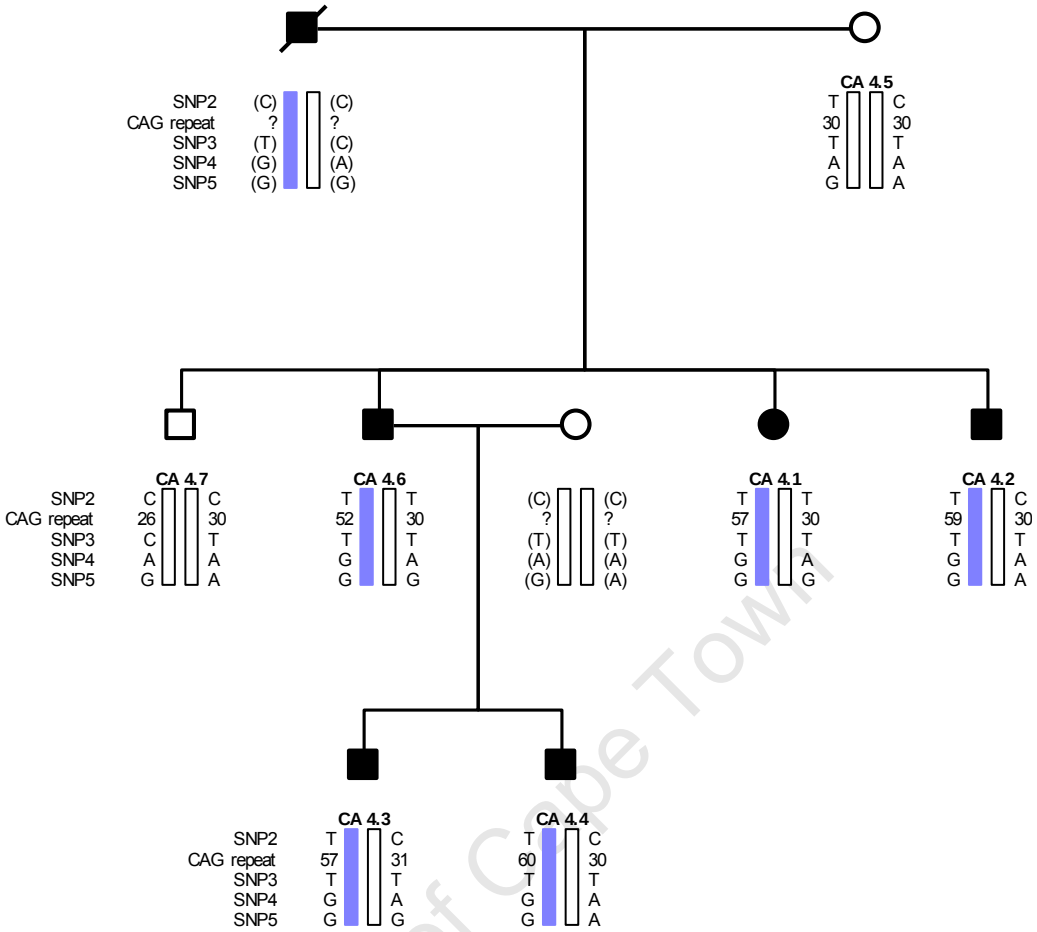


Figure 3.3 c) Pedigree from Family 4.

Haplotypes constructed with SNP genotypes obtained from this study. All loci are shown in 5' to 3' orientation. Blue bars represent the haplotype associated with the CAG expansion. Circles represent females, squares represent males. Filled in symbols are affected individuals. A line through the symbol represents a deceased individual. Genotypes in brackets "()" are inferred. "?" represents an unknown genotype.

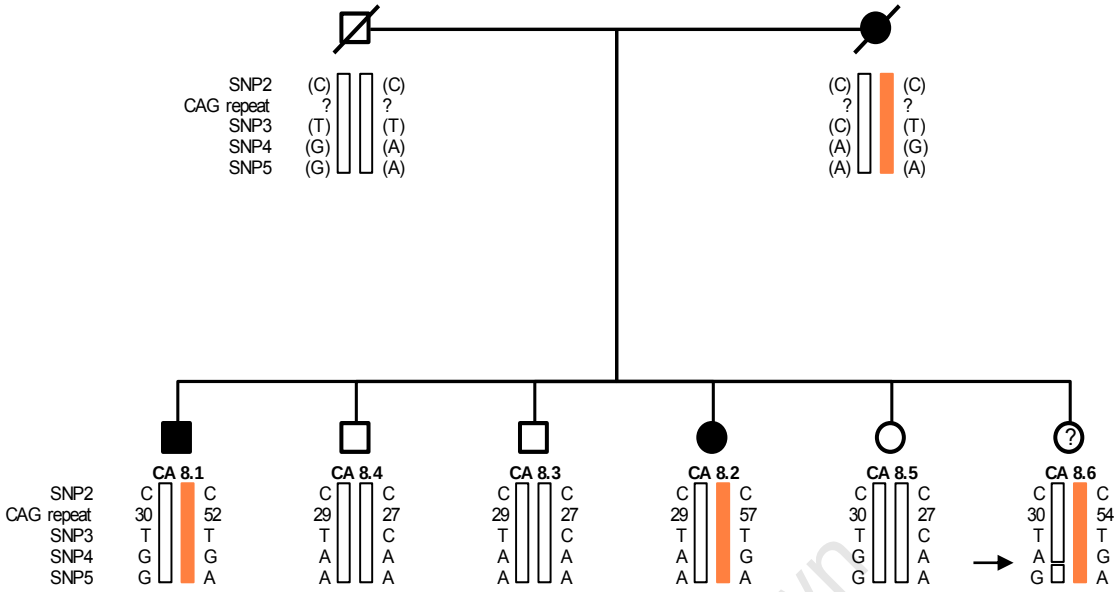


Figure 3.3 d) Pedigree of Family 8.

Haplotypes constructed with SNP genotypes obtained from this study. All loci are shown in 5' to 3' orientation. Orange bars represent the haplotype associated with the CAG expansion. Circles represent females, squares represent males. Filled in symbols are affected individuals. A line through the symbol represents deceased individuals. Symbols with "?" inside indicate unknown clinical status. Genotypes in brackets "()" are inferred. "?" represents an unknown genotype. → indicates a possible recombination event.

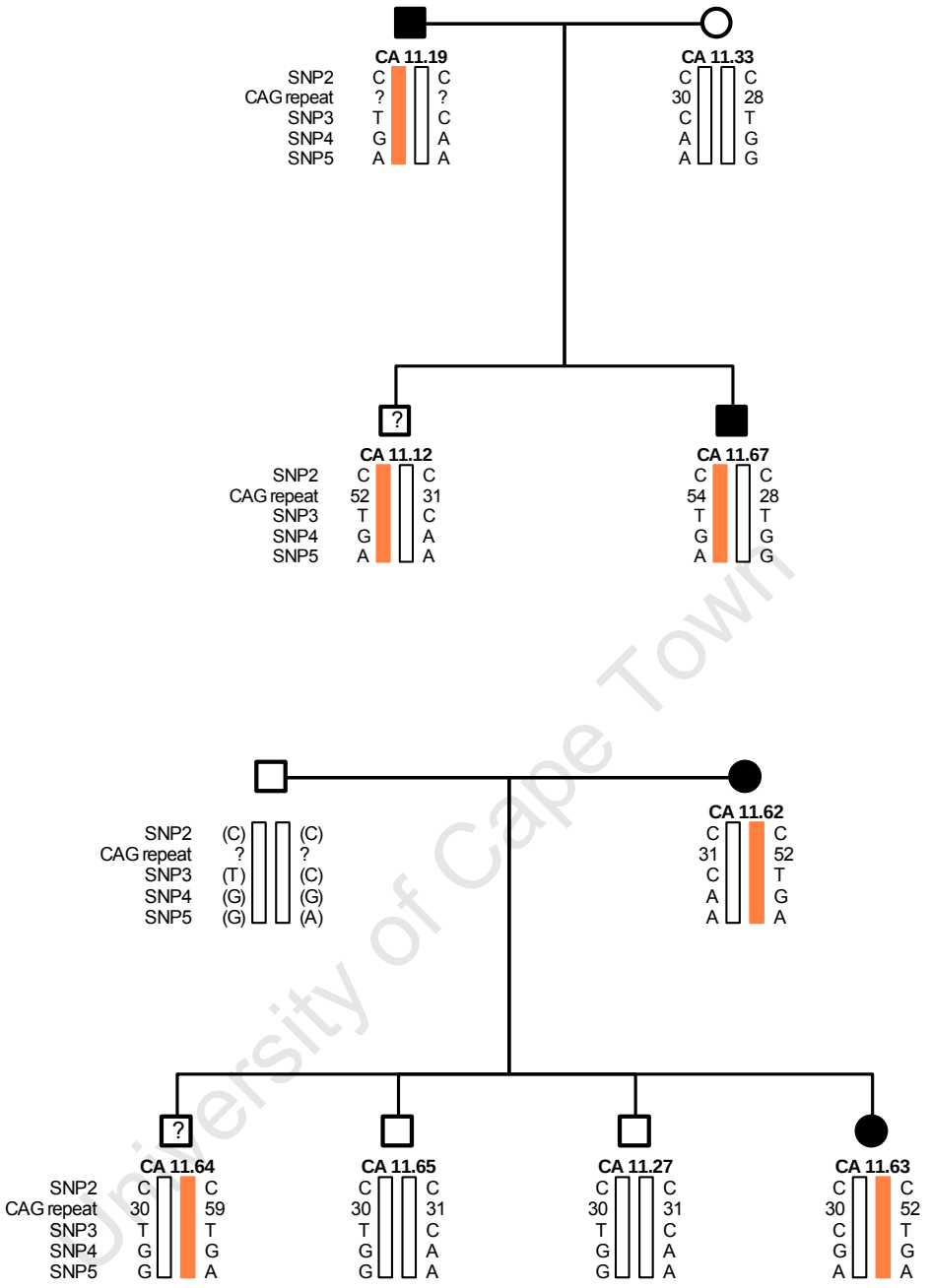


Figure 3.3 e) Pedigrees from Family 11.

Haplotypes constructed with SNP genotypes obtained from this study. All loci are shown in 5' to 3' orientation. Orange bars represent the haplotype associated with the CAG expansion. Loci with missing genotype data are represented by "?". Circles represent females, squares represent males. Filled in symbols are affected individuals. Symbols with "?" inside indicate unknown clinical status. Genotypes in brackets "()" are inferred. "?" represents an unknown genotype.

It should be noted that for an autosomal dominant disorder such as SCA1, only half of the alleles investigated are associated with the disease-causing mutation. Table 3.2 below summarises the two disease-associated haplotypes. The colours in the table correspond to those of the haplotypes linked to the expansion in the pedigrees (figure 3.3). The blue colour highlights families 1, 2 and 4 associated with haplotype 1 (Hap1) while orange highlights families 8 and 11 associated with haplotype 2 (Hap2). Also highlighted is a minimum shared interval common across both haplotypes (red box). This interval encompasses SNPs 3 and 4 which are downstream of the CAG repeat, 5' to 3' orientation). The two SNPs were in the previous section found to be closely linked to the CAG expansion in both disease-associated haplotypes.

Table 3.2 Mixed Ancestry family haplotypes associated with the SCA1 expansion.

SNP	*FAMILY*				
	CA 1	**CA 2**	**CA 4**	**CA 8**	**CA 11**
SNP2	T	T	T	C	C
Position of CAG repeat					
SNP3	T	T	T	T	T
SNP4	G	G	G	G	G
SNP5	G	G	G	A	A

Key:

= Minimum shared interval

blue background = Hap1, orange background = Hap2

Note that the haplotypes highlighted in the table were those constructed from the pedigrees. Indicated alleles are in the 5' to 3'direction (forward strand).

3.3 Statistical analysis

The SCA1 affected individuals (patients) in this section were all part of the family studies (Section 3.2). A total of 27 affected individuals and 75 unaffected unrelated controls were genotyped and analysed. The controls were analysed in order to establish HWE for the selected SNPs; and to test the occurrence of the disease-associated haplotypes in the background population using PHASE.

3.3.1 Hardy-Weinberg equilibrium

Genotyping data from 75 unrelated unaffected control subjects of Mixed Ancestry origin was used to establish which SNPs are in HWE in this population. This was done by performing an exact test (degree of freedom = 1) and assigning a corresponding p-value (significance). A p-value less than 0.05 is indicative of a population that is out of HWE. All 4 SNPs were found to be in HWE with p-values greater than 0.1.

3.3.2 PHASE analysis

PHASE analysis was performed on the control cohort to test the occurrence of the disease-associated haplotypes in the background population. PHASE was also used on the affected individuals to test the frequency of the disease-associated haplotypes obtained by segregation analysis. In order to prevent bias, separate input files were created for 27 SCA1 affected individuals from the 5 Mixed Ancestry families, and for 75 unaffected unrelated controls from the Mixed Ancestry sub-population. Analysis was performed on the two groups. The output files (frequencies) obtained from PHASE are shown in figure 3.4 below:

a)

index	haplotype	E(freq)	S.E
1	1101	0.080000	0.000000
2	**1100**	**0.320000**	**0.000000**
3	1001	0.080000	0.000000
4	1000	0.040000	0.000000
5	**0101**	**0.200000**	**0.000000**
6	0100	0.020000	0.000000
7	0111	0.060000	0.000000
8	0001	0.140000	0.000000
9	0011	0.060000	0.000000

b)

index	haplotype	E(freq)	S.E
1	0000	0.029407	0.011278
2	0001	0.057340	0.016142
3	0010	0.004305	0.004983
4	0011	0.062160	0.014061
5	0100	0.072459	0.014579
6	**0101**	**0.051153**	**0.015120**
7	0110	0.013479	0.006011
8	0111	0.026581	0.010474
9	1000	0.151584	0.017961
10	1001	0.182046	0.020321
11	1010	0.043998	0.012696
12	1011	0.080525	0.017928
13	**1100**	**0.041904**	**0.015776**
14	1101	0.085288	0.016829
15	1110	0.024698	0.011524
16	1111	0.073074	0.015894

Figure 3.4 Output files from PHASE analysis of Mixed Ancestry patients and controls. a) Output file from PHASE analysis of Mixed Ancestry patients. b) Output file from PHASE analysis of Mixed Ancestry controls. Nucleotides are represented by 0s and 1s (A/T = 0, G/C = 1). E(freq) is an estimate of the haplotype frequency of the sample. S.E is the estimated standard deviation for the sample. Highlighted in red are the haplotypes under investigation.

The estimated haplotype frequencies of each sample analysed can be used as estimates of the population haplotype frequencies. The two disease-associated haplotypes highlighted in red, 1100 and 0101 (figure 3.4), correspond to Hap1 and Hap2 respectively (table 3.2) in the following order: SNP5-SNP4-SNP3-SNP2. As expected, results from PHASE show that the two haplotypes have significantly higher frequencies in the affected individuals Hap1 E(freq)=0.32 and Hap2 E(freq)=0.20, compared to the controls Hap1 E(freq)=0.04 and Hap2 E(freq)=0.05.

SECTION II: Allele-specific silencing

3.4 Selected SNP targets

Subsequent to the genotyping of SNPs for the SNP-based haplotype, two SNPs were identified as possible targets for allele-specific silencing, with both showing fairly high heterozygosity. SNPs 2 and 4 were heterozygous in 62% and 55% respectively, of the patient population. However, it was determined that the mRNA sequence (NM_000332.3, NCBI) used as a reference in this study does not carry SNP2. Allele-specific silencing was thus attempted using effectors targeting SNP4 for which the G allele is associated with the expansion in both haplotypes identified in the Mixed Ancestry SCA1 families.

3.5 Plasmid construction/cloning

3.5.1 shRNA cassettes

Subsequent to successful PCR, the U6/shRNA cassettes were cloned into the pGEM-T Easy vector. Plasmid DNA was purified and positive clones selected by restriction enzyme digestion with *Eco*RI. DNA sequencing was performed to confirm the correct shRNA sequence for each construct. Large scale plasmid purification was then performed to yield sufficient pDNA for transfection into HEK293 cells.

3.5.2 Targets

The short *ATXN1* target sequences were successfully cloned into the psiCheck 2.2 vector, downstream of the *Renilla* reporter gene. Plasmid DNA was purified and positive clones selected by restriction enzyme digest with *Eco*RV. DNA sequencing was performed to confirm the correct target sequence. Large scale plasmid purification was then performed to yield pDNA for transfection.

3.6 Transfection studies

A GFP-expressing plasmid was co-transfected with the shRNA- and target-expressing plasmids. Twenty four hours after transfection, the cells were viewed under a fluorescent microscope and photographed. GFP expression gives an indication of efficiency (figure 3.5). Cells in figure 3.5 a) were approximately 80% confluent and clearly show a fairly high level of transfection.

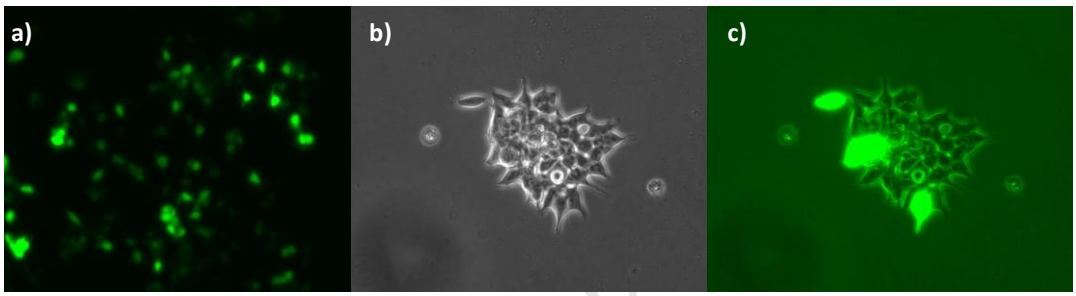


Figure 3.5 Representative images of transfected HEK293 cells viewed under a fluorescent microscope. a) Cells under green fluorescence (10X). b) Cluster of cells under brightfield (20X). c) Merged images of cells in (b) under green fluorescence and brightfield (20X).

3.7 Results of a dual luciferase reporter assay

Multiple shRNAs were designed to incorporate the mismatch created against the wild-type transcript. The mismatch was placed successively in positions 9 to 16 from the 5' end of the guide strand. Forty eight hours after transfection, a dual luciferase reporter assay was used to screen the shRNAs against short target sequences (62 nt) of either the wild-type or the mutant transcript. The results of this assay are displayed graphically in figure 3.6 below.

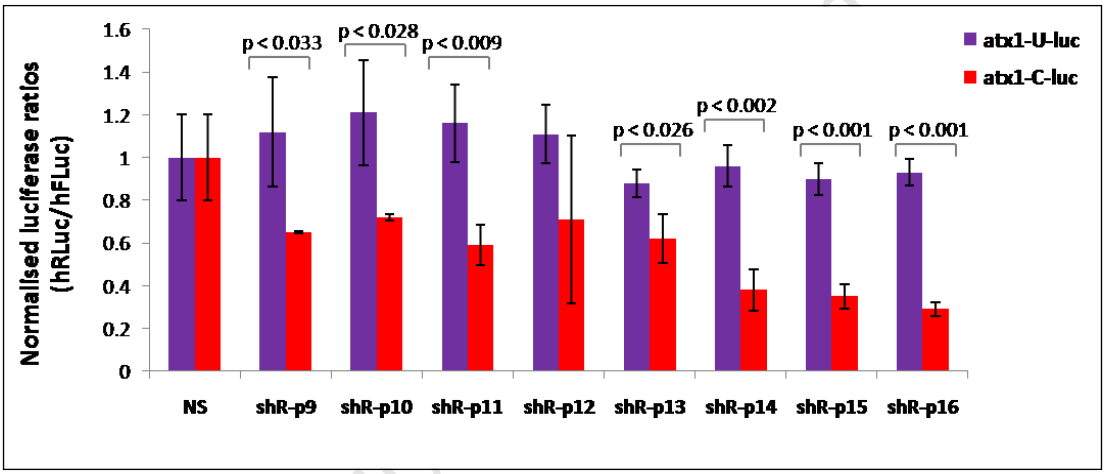


Figure 3.6 A graph showing the results of a dual luciferase assay using a series of mismatched shRNAs targeting SNP4. The relative values of *Renilla* luciferase (hRLuc) expression are normalised to firefly luciferase (hFLuc) expression in each case. Each experiment was performed in triplicate. Data was averaged and normalised to the non-specific (NS) control. Relative expression of wild-type (atx1-U-luc) and mutant (atx1-C-luc) targets is represented by purple and red bars respectively with error bars showing standard deviation. Statistically significant differences ($p < 0.05$) between wild-type and mutant knockdown are indicated.

Other than shR-p12, all the constructs showed statistically significant discrimination between the wild-type and mutant ($p < 0.05$) as indicated in figure 3.6. Results from shR-p12 showed a p-value which is not statistically significant. This was due to an outlier value in the measured luciferase activity which creates a very large standard deviation. Results from the other constructs show a trend in discrimination with an increase from left to right of the figure i.e. the further away the mismatch is from the seed region. shR-p9, -p10 and -p11 show significant discrimination but also have fairly wide error bars. shR-p13 shows knockdown of the wild-type and the mutant to approximately 90% and 60% respectively. shR-p14 shows minimal knockdown of the wild-type to just under 100%, while the mutant is reduced to just below 40%. shR-p15 shows very significant discrimination ($p<0.001$) with knockdown of wild-type and mutant at approximately 90% and 30% respectively. shR-p16 is the most effective construct with minimal knockdown of the wild-type to approximately 90% and the mutant to just under 30%, relative to the NS control.

4. Discussion

4.1 Haplotype study

Due to numerous large studies carried out only in first world countries, there is a gross under-representation of African genotypes on international databases such as NCBI. Examination of the region of interest for SNPs to be used in the construction of the SNP-based haplotype reinforced this. There is a severe lack of genotype information specifically for the sub-Saharan African populations, and several SNPs were disqualified due to the unavailability of minor allele frequencies and heterozygosity information. However, the HapMap (www.hapmap.org), 1000 Genomes (www.1000genomes.org) and the Africa Genome (www.africagenome.com) projects are attempting to address the issue of a lack of collated information by genotyping various African populations.

4.1.1 Confirmation of reported haplotypes

The first aim of the present study was to narrow the region investigated by Ramesar and colleagues (Ramesar *et al.* 1997) and to confirm the two founder effects by constructing a SNP-based haplotype. The SNP-based haplotype was successfully constructed using genotypes from four SNPs across a region spanning the *ATXN1* gene.

Figure 4.1 shows the positions of the *ATXN1* gene, the SNPs used in this study and the microsatellite markers from the previous study.

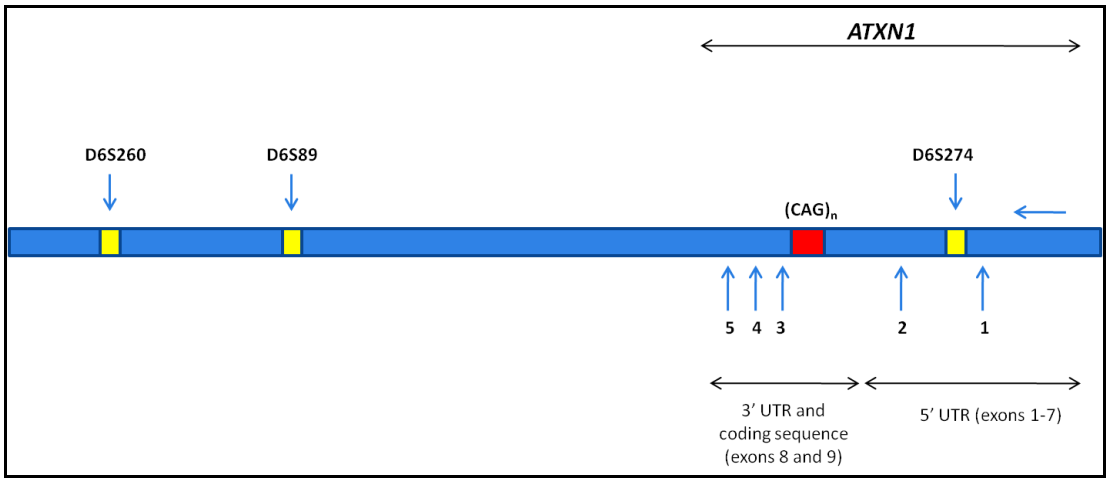


Figure 4.1 Diagram showing the *ATXN1* gene, and the positions of the microsatellite markers (D6S260, D6S89 and D6S274) in yellow (Ramesar *et al*. 1997). 1, 2, 3, 4 and 5 indicate the positions of the SNPs used in the construction of the haplotype in this study

The previous study investigated a large genomic area (approximately 10 Mb), as shown by the positions of the microsatellite markers in figure 4.1. In this study, the region investigated was approximately 470 kb. By narrowing the region using SNPs, the haplotypes constructed in this study have confirmed the two founder effects in the SCA1 families of Mixed Ancestry origin in the Western Cape. Families 1, 2 and 4 share a common haplotype designated Haplotype1 (Hap1), and families 8 and 11 share a second common haplotype designated Haplotype2 (Hap2) (table 3.2). These results correspond with the haplotypes assigned to the families in the previous study.

As mentioned briefly in chapter 3, SNP1 (rs471716) showed the same genotype (GG) for every individual. This SNP is located 5' to the microsatellite marker DS6S274 (in the direction of transcription) (figure 4.1). Further analysis of this region is required to ascertain the extent of the haplotype and breakpoints of the two haplotypes. This may be done by selecting and genotyping additional SNPs 5' to the marker D6S274 (i.e. between the marker and SNP1) in order to determine the extent of the haplotype.

The results of such an analysis could provide more information related to the origins of the haplotypes in the Mixed Ancestry population.

Table 3.2 summarises the haplotype information from the segregation analysis performed in this study. The table shows the existence of a minimum shared interval common across both haplotypes (red box). The interval encompasses SNPs 3 and 4 (rs2075974 and rs180017 respectively), indicating that these two SNPs are in close linkage disequilibrium with the CAG expansion. This may suggest the possibility of a common origin for the two founder haplotypes, and thus the existence of a single original SCA1 haplotype in these Mixed Ancestry families.

4.1.2 A different SNP haplotype identified in Family 8

The SNP haplotypes constructed for individual CA8.6 in family 8 suggest the possibility of a recombination event on the normal allele. In figure 3.3 (d), the break-point in the haplotype is indicated by (➞). Due to the unavailability of DNA samples, parental genotypes were inferred from the haplotypes constructed for the children. A cross-over event occurring between the normal alleles of the father, in the region between SNPs 4 and 5, may account for the different haplotype (C-T-A-G) in CA8.6.

Alternatively, the haplotype may have a different paternal source. It is note-worthy that this haplotype (C-T-A-G) in individual CA8.6 corresponds to 1001 in the output file from PHASE analysis of the controls (highlighted in green, figure 4.2 below).

index	haplotype	E(freq)	S.E
1	0000	0.029407	0.011278
2	0001	0.057340	0.016142
3	0010	0.004305	0.004983
4	0011	0.062160	0.014061
5	0100	0.072459	0.014579
6	**0101**	**0.051153**	**0.015120**
7	0110	0.013479	0.006011
8	0111	0.026581	0.010474
9	1000	0.151584	0.017961
10	**1001**	**0.182046**	**0.020321**
11	1010	0.043998	0.012696
12	1011	0.080525	0.017928
13	**1100**	**0.041904**	**0.015776**
14	1101	0.085288	0.016829
15	1110	0.024698	0.011524
16	1111	0.073074	0.015894

Figure 4.2 PHASE output (controls). Disease-associated haplotypes are highlighted in **red**, the different haplotype from Family 8 is highlighted in **green**.

The E(freq) for 1001 is 0.18 in the background population. This is the highest value for the SNP haplotypes constructed randomly by PHASE. The fact that this is the most common haplotype in the background population thus lends support that the paternal haplotype of individual CA 8.6 may have an alternative source. Unfortunately, due to the lack of DNA from the parents, there is currently no way of ascertaining which of these possibilities explains the occurrence of this haplotype.

4.1.3 SCA1 founder effects in the Mixed Ancestry families

The existence of a founder effect is not unusual in South Africa and has been previously reported for various genetic conditions such as keratolytic winter erythema (KWE) (Starfield *et al.* 1997), hypertrophic cardiomyopathy (HCM) (Moolman-Smook *et al.* 1999), SCA7 (Greenberg *et al.* 2006) and HD (Scholefield and Greenberg 2007). A founder effect may be described as the loss of genetic variation. This usually occurs when a small group of individuals with origins in a larger group, establishes a new population in which the frequency of certain genetic variations and mutations increases due to limitations in the gene pool.

The South African population is made up of groups of varying ethnic origin as a result of its location and colonisation history (de Wit *et al.* 2010). The main sub-populations are the Black African, Caucasian and Mixed Ancestry. The Mixed Ancestry population makes up just over 50% of the population of the Western Cape and is characterised by admixture of different populations. Recent genetic studies have supported historical data, showing the main components of the Mixed Ancestry population to be the indigenous African people (the Khoesan and Bantu-speaking Africans), Europeans and peoples from Malaysia, Indonesia and the Indian subcontinent (Moolman-Smook *et al.* 1999, Quintana-Murci *et al.* 2010, de Wit *et al.* 2010).

A study performed in India reported the existence of a SCA1 haplotype in that population and linked rs2075974 (SNP3 in this study), to large normal (LN) alleles and expanded alleles (Mittal *et al.* 2005). The authors reported the T allele of this SNP linked to CAG repeats with two CAT interruptions. As discussed in chapter 1, these interruptions are believed to stabilise the CAG repeat.

The present study identified the T allele of this SNP in linkage disequilibrium with the expansion in SCA1 affected individuals, but made no assessment of interruptions. The South African population includes people of Indian origin and it may be relevant that the same SNP is linked to the expansion in both populations. Extending the haplotype constructed here to include the other SNPs used in the study of the Indian population may yield information with regard to this. It may also be useful to investigate the occurrence of CAT interruptions in a SCA1 cohort. Analysis of the CAT interruptions is performed by restriction enzyme digestion using *Sfa*NI (Zühlke *et al.* 2002). This information may then be extrapolated to the diagnostic service offered by the Human Genetics Division, UCT.

An attempt was also made to extend the haplotype to other ethnic groups by genotyping other SCA1 affected individuals. It was hoped that computational analysis following the genotyping, might be used to determine the origin of the SCA1 mutation. Unfortunately it proved to be impossible to perform this analysis essentially due to a lack of information in patient records, specifically on ethnicity. South Africa has had a difficult history with discrimination based on race. It is still a sensitive issue today and studies such as this are hindered by the lack of information that results from the perceived discriminatory nature of questions pertaining to race and ethnic background. However, it is now better understood that this information is relevant to genetic studies. Extending the haplotype to other ethnic groups may become possible as more patients and families are added to the Human Genetics Division database, which may provide sufficient numbers for analysis.

The occurrence of common haplotypes in the SCA1 patients (at least of Mixed Ancestry) in South Africa, suggests that the form of therapy investigated here may well be worth exploring further as it would benefit a larger percentage of patients than in populations where there is no common disease-associated haplotype.

4.2 Allele-specific silencing

Silencing of dominant genes has been shown to be useful in the development of therapy for various diseases such as ALS (Ding *et al.* 2003), HD (Harper *et al.* 2005) and SCA3 (Miller *et al.* 2003). Although a SCA1 knock-out mouse model shows only a mild phenotype, allele-specific silencing is preferable for SCA1 in light of the fact that the exact function of wild-type ataxin-1 is unknown and its absence has been shown to have some neurobehavioral effects (Matilla *et al.* 1998, Xia *et al.* 2004).

The second aim of this study was to identify any appropriate single nucleotide differences between wild-type and mutant SCA1 gene transcripts that may be exploited for allele-specific RNAi as has been done previously for SCA3 (Miller *et al.* 2003, Alves *et al.* 2008), and more recently for HD (Zhang *et al.* 2009) and SCA7 (Scholefield *et al.* 2009). The results of this current study show that SNP4 (rs180017) may be successfully targeted to achieve allele-specific silencing for short target sequences of the SCA1 mutant gene transcript.

4.2.1 Target selection

The attempt at allele-specific silencing followed on Section I of this study in which a SNP-based haplotype was constructed. Subsequent to genotyping of the SNPs selected for construction of the haplotype, two SNPs were identified in as possible targets for allele-specific silencing of the SCA1 gene. The two SNPs, SNP2 (in the 5' UTR) and SNP4 (in the 3' UTR) were found to be heterozygous in 62% and 55% respectively of the affected individuals genotyped in this study. This level of heterozygosity initially suggested that successful targeting of either or both of these polymorphisms would make a significant difference to this patient population.

Two online databases (Ensembl and NCBI) were used to select SNPs and obtain information on their genomic positions, heterozygosity and MAF. To investigate silencing of the mutant transcript, it was then necessary to identify the mRNA transcripts of the *ATXN1* gene to determine the positions of the two SNPs considered as targets. As of November 2008, the Ensembl database had 2 different transcripts of the ATXN1 gene, one with 11 exons with both proposed target SNPs, and the other with 9 exons. The NCBI database had 2 variants of the same transcript (1 and 2), neither of which had SNP2. Subsequent to comparison of these transcripts and consultations with the database managers, it was determined that due to alternative splicing in intron 7 SNP2 was absent from the coding region of the gene. It was thus decided that variant 1 from NCBI (NM_000332.3) would be used as the reference sequence for this study. Ensembl database managers consequently retracted the transcript with 11 exons that included SNP2 in the coding sequence, and as of December 2009 published a single transcript with 9 exons. Therefore, SNP4 (rs180017), for which the G allele was found to be linked to the disease-causing expansion in over half of the South African SCA1 patients analysed, was selected as the target SNP.

Effectors were designed to target this SNP in order to test the feasibility of allele-specific silencing of the mutant transcript in this cohort of SCA1 patients. Figure 4.3 shows how allele-specific silencing may be achieved by targeting SNP4.

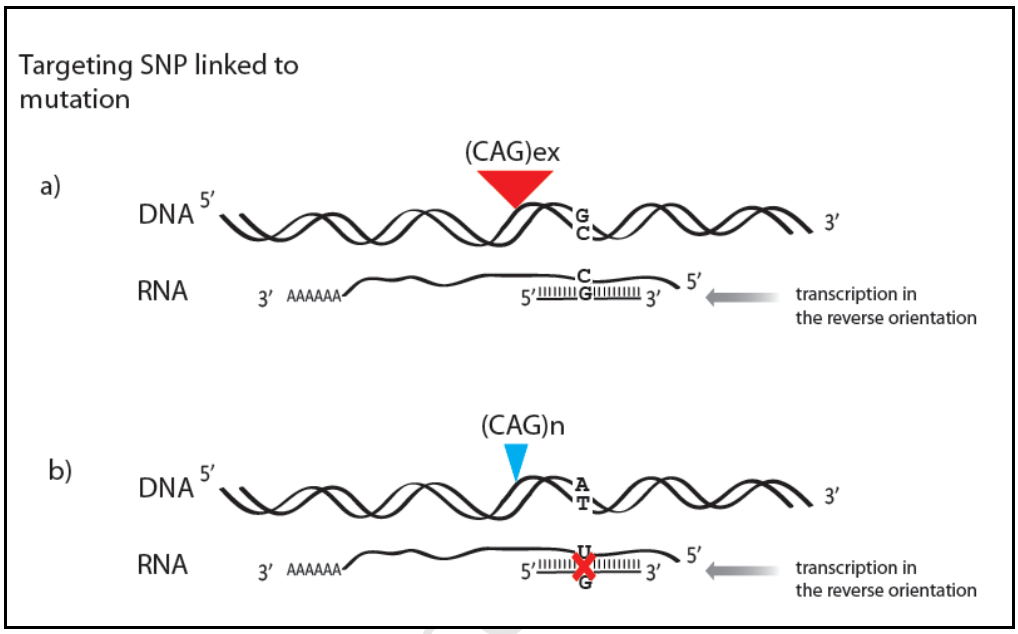


Figure 4.3 Allele-specific silencing may be achieved by targeting a heterozygous SNP. a) An effector sequence perfectly complementary (G:C) to the mutant transcript leads to its cleavage. b) The same effector sequence with incomplete complementarity (G:U) to the wild-type transcript does not lead to cleavage and allows expression of the wild type protein.

The G allele (forward genomic sequence) linked to the expansion, is transcribed to a C in the mRNA sequence. Therefore effectors were designed incorporating a G in this position to complement the mutant transcript and introduce a mismatch to the wild-type transcript (figure 4.3). The wild-type genomic allele is an A which is transcribed to a U in the mRNA sequence.

The mismatch introduced between the effector sequences and the mRNA transcript is a G:U, commonly referred to as a G:U wobble. This is a weak mismatch due to the formation of a single hydrogen bond between the purine G and pyrimidine U (Schwarz *et al.* 2006).

4.2.2 Mismatch placement

Several factors must be taken into consideration when attempting allele-specific silencing; the strength of the mismatch, its position within the guide sequence and whether or not the target sequence is amenable for RNAi. The type (and therefore strength) of the mismatch and accessibility of the target cannot be controlled since these depend on the SNP itself. The placement of the mismatch however can be varied and this has been shown in several studies (Miller *et al.* 2003, Ding *et al.* 2003, Li *et al.* 2004, Schwarz *et al.* 2006).

It was previously believed that placing the mismatch between position 10 and 11 (with position 1 being the first nt of the 5' end of the guide strand), would achieve the greatest discrimination by impairing cleavage of target mRNA at its natural site and thus minimising knock-down of the wild-type allele (Leuschner *et al.* 2006). However, an extensive study performed by Schwarz and colleagues in 2006, reported that mismatches 3' to the seed region promoted single nucleotide discrimination. Short interfering RNAs with the mismatch in positions 10 and in particular 16 consistently showed high levels of discrimination (Schwarz *et al.* 2006). A similar result was obtained by Scholefield *et al.* targeting the *atxn7* mutant transcript (Scholefield *et al.* 2009) with full-length assays showing the mismatch in position 16 achieving the greatest selectivity. Thus for this study, a set of eight effector sequences were designed with the mismatch placed from position 9 to 16.

4.2.3 Dual luciferase reporter assay

A dual luciferase reporter assay was used to investigate the feasibility of targeting the particular SNP in this sequence as well as the placement of the mismatch. The results of the luciferase assay (figure 3.6) demonstrate that despite using a fairly weak mismatch to the wild-type allele (figure 4.3), there is significant discrimination shown at almost every position. Positions 14, 15 and 16 show a particularly high level of discrimination. Using tiled siRNAs Schwarz and colleagues reported that little or no discrimination was achieved by mismatches other than of purine:purine siRNA:target RNA pairing (Schwarz *et al.* 2006). Our findings are in contrast to this, with fairly substantial discrimination being achieved by a purine: pyrimidine mismatch.

To our knowledge, this is the first study that investigates allele-specific silencing for SCA1 and only the second study to successfully show discrimination between alleles based on a weak mismatch. As such, this work supports the development of allele-specific silencing for SCA1 patients in South Africa. It also supports the use of this approach for other disorders and conditions with discrimination based on weak mismatches. The findings of this study are also in line with the results reported by Schwarz *et al.* and Scholefield *et al.*, with the mismatch at position 16 in the hairpin achieving the highest level of discrimination in this system (Schwarz *et al.* 2006, Scholefield *et al.* 2009). In light of the positive results described here, there is much more to be done to investigate the development of allele-specific silencing for the SCA1 gene. Some suggestions are outlined in the next section addressing future work. Any information derived from the development of this technique as a form of therapy may well be applied to other population groups around the world to alleviate disease symptoms in these individuals and relieve the burden of SCA1.

5. Future prospects and conclusion

5.1 SCA1 haplotypes

Further analysis is required with regard to the SNP-haplotypes constructed. Unfortunately, this study was unable to determine the extent of the haplotypes and therefore the origins. To this end, extending the SNP-based haplotype to SCA1 individuals from other ethnic groups in South Africa may provide additional information.

As mentioned, an investigation into the occurrence of CAT interruptions in SCA1 affected individuals may also provide some information relevant to the origin of the founder haplotypes identified. In spite of the fact that CAT interruptions have been reported to stabilise the CAG repeat, this analysis is not offered as part of the molecular diagnostic service in the Human Genetic Division at UCT. A study into the frequency of interruptions in these SCA1 individuals/families may add significance to diagnostic results. Individuals carrying a CAG repeat with a stabilising interruption are unlikely to transmit an expanded repeat to their offspring and this could have an impact on reproductive choices.

5.2 Development of an RNAi-based therapy for SCA1 in South Africa

5.2.1 Cellular models

Given the success realized by targeting this sequence, future work should focus on whether these results can be replicated in a system that more closely mimics the SCA1 disease state such as in a neuronal cell line or a SCA1 mouse model. As investigated here, using short targets around the selected SNP does not give complete information with regard to the knock down of the target mRNA.

It has been reported that the secondary structure of full-length mRNA can affect the efficiency of cleavage by RISC (Shao *et al.* 2007). The silencing of SCA7 investigated by Scholefield *et al.* supported this, with knock down efficiency of full-length targets lower than that achieved in a luciferase assay (Scholefield *et al.* 2009), and the most discriminating effector having the mismatch in different positions in the two systems. The shRNAs designed in this study thus need to be screened against full-length *ATXN1* transcripts fused to reporter genes in a heterozygous cellular system. In the heterozygous system the mutant and wild-type targets are co-transfected, giving a closer approximation of the *in vivo* state since both mutant and wild-type transcripts exist together within the patient cells.

This study and many others use human embryonic kidney (HEK) cells due to their amenability for cell culture and transfection; however they are generally not considered an ideal model. The inability to sample live brain cells severely limits knowledge of human neuropathology (Marchetto *et al.* 2010).

A closer cellular model for a neurodegenerative disease such as SCA1 would thus be beneficial. In spite of the wide-spread expression of ataxin-1 protein, disease pathology is only evident in a small subset of neuronal cells and the reason for this selectivity is not yet understood. Recently, advances have been made to enable a study that gives a more ideal model of disease using induced pluripotent stem (iPS) cells (Chamberlain *et al.* 2008, Marchetto *et al.* 2010).

Human iPS cells are defined as having the potential to give rise to many, or all human tissue types. The cells can be generated by introducing reprogramming factors using retroviral vectors. These factors have the ability to reverse a differentiated cell to a pluripotent state, such that the cell can then be driven to form a different cell type. This technique thus offers the opportunity to model specific disease from available patient fibroblasts (Chamberlain *et al.* 2008, Vierbuchen *et al.* 2010, Marchetto *et al.* 2010). Ebert and colleagues reported the generation of iPS cells from skin fibroblasts taken from a patient with spinal muscular atrophy (SMA), the first study to show that human iPS cells can be used to model the specific pathology seen in an inherited disease (Ebert *et al.* 2009). Several other studies have also examined the reprogramming of fibroblasts for human neurological diseases such as amyotrophic lateral sclerosis (ALS) and Parkinson disease (PD) with some success (Marchetto *et al.* 2010). There is much work to be done yet, but iPS cells present enormous potential both as models of disease as well as for drug discovery and the testing of therapeutics.

5.2.2 Modification of effectors

Experiments using these guide sequences in pri-miRNA would be a positive step towards the development of an RNAi-based therapy for South African SCA1 patients specifically, as well as other populations around the world. In comparison with the shRNA molecules used in this study, pri-miRNAs follow a more natural pathway since they closely resemble the endogenous RNAi substrates (Boudreau *et al.* 2008, Weinberg and Wood 2009). They also have the advantage of tissue-controlled expression, whereas the shRNA expression off Pol II or Pol III promoters for the most part cannot be tightly regulated which may have undesirable effects on the RNA processing pathway.
As mentioned in chapter 1 (section 1.5), the development of a miRNA expression system may circumvent the issues of neurotoxicity as reported in a knockin HD mouse model (McBride *et al.* 2008).

In 2008, Ohnishi and colleagues reported improved discrimination between wild-type and mutant alleles by the introduction of base substitutions in the guide strand of siRNAs and shRNAs (Ohnishi *et al.* 2008). The resulting sequences are predicted to have one mismatch against the mutant and two against the wild-type allele. The authors speculate that the disruption of base-pairing in the seed region as well as the centre of the guide strand reduces recognition and/or silencing against the wild-type alleles possibly due to the thermodynamics involved (Ohnishi *et al.* 2008). However, supplementary data from the study by Scholefield and colleagues suggests that the use of a double mismatch to the wild-type does not significantly improve discrimination (Scholefield *et al.* 2009). It may still be worthwhile to examine the effect of such modifications on the effector sequences investigated here.

A major issue that must yet be addressed is delivery to the CNS. As outlined in chapter 1, there are various hurdles to delivering RNAi effector molecules to the brain. Several studies have had success with various modifications of the molecules as well as delivery of expression vectors (Singh and Hajeri 2009), and investigation is on-going into safe and specific delivery of RNAi effectors.

5.2.3 Animal models

RNAi as a therapeutic technique has already been tested in a SCA1 transgenic mouse model with promising results (Xia *et al.* 2004). Xia and colleagues used recombinant AAV vectors expressing shRNAs flanking the CAG repeat region. The vectors were injected into the cerebellums of the SCA1 mice and results showed a marked improvement in cellular and behavioural characteristics. A similar model needs to be used to test the efficacy as well as required dosage of the effectors and at what stage the therapy should be applied. Another study on a conditional SCA1 mouse model suggests that early removal of the mutant protein is beneficial (Zu *et al.* 2004).

Various investigations into animal models also suggest that allele-specific RNAi may not be essential for SCA1 therapy. Ataxin-1 knockout mice show only mild neurobehavioral characteristics (Matilla *et al.* 1998). Thus, it may not be necessary to achieve complete selectivity between mutant and wild-type alleles in order to offer an RNAi therapeutic. If the mutant allele is knocked down to a large extent, then it is possible that a fairly low level of wild-type may be sufficient for normal neurological functioning.

This kind of non allele-specific silencing has recently been investigated in rodent models for HD (Drouet *et al.* 2009, Boudreau *et al.* 2009). Lentiviral vectors were developed to allow the regulated expression of shRNAs targeting either wild-type or mutant alleles. Results suggest that local inactivation of endogenous huntingtin in brains of adult mice is feasible and has no serious consequences, at least over several months. Residual protein levels appear to be sufficient for biological functioning.

Another study has taken a similar approach for SCA3. Alves *et al.* used LV vectors to either over-express wild-type human ataxin-3 or silence endogenous ataxin-3 in a rat model of SCA3 (Alves *et al.* 2010). Results indicated the absence of a role for ataxin-3 and led them to investigate the silencing of both mutant and wild-type ataxin-3. Non-allele-specific silencing led to robust reduction of neurodegeneration (Alves *et al.* 2010). For both HD and SCA3 global silencing appears to offer the chance to treat all patients with one product. This kind of approach may be well-worth investigating in the development of a therapy for SCA1 but for any of these diseases safety and efficacy need to be extensively studied in different disease models. Alternatively, global knockdown of the mutant and wild-type transcripts and simultaneous replacement of the wild-type with an RNAi- resistant transgene may be possible (Scholefield and Wood 2010). The choice of which approach to employ is largely dependent on how tolerable loss of wild-type function is.

5.3 Conclusion

In conclusion, further analysis is required with regard to the SCA1 founder haplotype in the South African Mixed Ancestry population. Extension of the haplotype to include other SNPs and investigation of patients from other ethnic groups may resolve the issue of the extent of the founder haplotype as well as its origins.

With regard to the attempt to target SNP4 (rs180017) for allele-specific silencing of the SCA1 gene, the positive results recorded here suggest that it may be possible to develop an RNAi-based therapy for the sub-set of South African patients investigated, and supports the use of this approach for other populations with similar founder effects.

Appendix A

Primer sequences

SCA1 PCR	**F** 5' gcggtcccaaaagggtcagt AAC TGG AAA TGT GGA CGT AC 3' ***R** 5' ggtcccaaaagggtcagt CAA CAT GGG CAG TCT GAG 3' primers obtained from the molecular diagnostic laboratory, Human Genetics Division (UCT); published sequences (Dorschner *et al*. 2002)
Single step plasmid PCR (construction of U6/shRNA cassettes)	**U6+1 universal F primer:** 5' GATCGCAGATCTGGATCCAAGGTCGGGCAGGAAGAGGGCCT 3' **Long reverse oligos:** **P9** 5'AAAAAGTGGGAAGGGCCCTGCAGGTGGGTCAGGCCCACAGGGCCCTTCCCACCGGTGTTTCGTCCTTTCCACAA 3' **P10** 5'AAAAATGGGAAGGGCCCTGCAGGGTGGGTCAGGCCCCACAGGGCCCTTCCCACGGTGTTTCGTCCTTTCCACAA 3' **P11** 5'AAAAAGGGAAGGGCCCTGCAGGGGTGGGTCAGGCCCCCACAGGGCCCTTCCCCGGTGTTTCGTCCTTTCCACAA 3' **P12** 5'AAAAAGGAAGGGCCCTGCAGGGGTTGGGTCAGGACCCCCACAGGGCCCTTCCCGGTGTTTCGTCCTTTCCACAA 3' **P13** 5'AAAAAGAAGGGCCCTGCAGGGGTGTGGGTCAGGCACCCCCACAGGGCCCTTCCGGTGTTTCGTCCTTTCCACAA 3' **P14** 5'AAAAAAAGGGCCCTGCAGGGGTGCTGGGTCAGGACACCCCCACAGGGCCCTTCGGTGTTTCGTCCTTTCCACAA 3' **P15** 5'AAAAAAGGGCCCTGCAGGGGTGCATGGGTCAGGCACACCCCCACAGGGCCCTCGGTGTTTCGTCCTTTCCACAA 3' **P16** 5'AAAAAGGGCCCTGCAGGGGTGCAGTGGGTCAGGCCACACCCCCACAGGGCCCCGGTGTTTCGTCCTTTCCACAA 3'
Sequencing of plasmid/insert (shRNAs) ‡	**SP6 primer**: 5' ATTTAGGTGACACTATAGAA 3' **T7 primer**: 5' TAATACGACTCACTATAGGG 3'
Sequencing of plasmid/insert (targets) ‡	**Psi-F:** 5' GACGCTCCAGATGAAATGGG 3' **Psi-R:** 5' GTGCCCGTGGCCACCAAGAC 3'

*****SCA1 R is tagged with 6-FAM, a fluorescent label.
Long R oligos were obtained from Integrated DNA Technologies (*Whitehead Scientific*).
The U6+1 universal F primer was provided by J. Scholefield.
SP6 and T7 primers were available in the Human Genetics Division, UCT.
Psi-F and Psi-F primers were a kind gift from M. Weinberg.

Design of oligos for shRNA construction

shR p9

5′ TTGTGGAAAGGACGAAACACCG GTGGGAAGGGCCCTGTGGG CCTGACCCA CTGCAGGGCCCTTCCCACTTTTT 3′
U6 snRNA promoter sequence **P** **miR-122 loop** **G**

shR p10

5′ TTGTGGAAAGGACGAAACACCG TGGGAAGGGCCCTGTGGGG CCTGACCCA CCCTGCAGGGCCCTTCCCATTTTT 3′

shR p11

5′ TTGTGGAAAGGACGAAACACCG GGGAAGGGCCCTGTGGGGG CCTGACCCA CCCCTGCAGGGCCCTTCCCTTTTT 3′

shR p12

5′ TTGTGGAAAGGACGAAACACCG GGAAGGGCCCTGTGGGGGT CCTGACCCA ACCCCTGCAGGGCCCTTCCTTTTT 3′

shR p13

5′ TTGTGGAAAGGACGAAACACCG GAAGGGCCCTGTGGGGGTG CCTGACCCA CACCCCTGCAGGGCCCTTCTTTTT 3′

shR p14

5′ TTGTGGAAAGGACGAAACACCG AAGGGCCCTGTGGGGGTGT CCTGACCCA GCACCCCTGCAGGGCCCTTTTTTT 3′

shR p15

5′ TTGTGGAAAGGACGAAACACCG AGGGCCCTGTGGGGGTGTG CCTGACCCA TGCACCCCTGCAGGGCCCTTTTTT 3′

shR p16

5′ TTGTGGAAAGGACGAAACACCG GGGCCCTGTGGGGGTGTGG CCTGACCCA CTGCACCCCTGCAGGGCCCTTTTT 3′

P – passenger/sense strand, **G** – guide/antisense strand

DNA oligonucleotides used for cloning of target plasmids (62 nt)

Wild-type target A:

5′ TCGAGATATCTCAGGAGAAAAGTCCCGTGGGAAGGGTCCTGCAGGGGTGCAGGGTTGCACGC 3′

Wild-type target B:

5′ GGCCGCGTGCAACCCTGCACCCCTGCAGGACCCTTCCCACGGGACTTTTCTCCTGAGATATC 3′

Mutant target A:

5′ TCGAGATATCTCAGGAGAAAAGTCCCGTGGGAAGGGCCCTGCAGGGGTGCAGGGTTGCACGC 3′

Mutant target B:

5′ GGCCGCGTGCAACCCTGCACCCCTGCAGGGCCCTTCCCACGGGACTTTTCTCCTGAGATATC 3′

Pink – *Xho*I recognition site, Purple – *Eco*RV recognition site, Green – *Not*I recognition site, Red – target SNP

Appendix B: General recipes

Tris Borate-EDTA (TBE) buffer 10X

- 108 g tris (hydroxymethyl) aminomethane
- 55 g boric acid
- 7.4 g ethylenediaminetetra-acetic acid (EDTA)

Dissolve and make up to a final volume of 1 litre with distilled water. Dilute as required.

Low electro-endosmosis (LE) agarose gels (w/v)

0.8%: 0.4 g LE agarose in a final volume of 50 ml 1X TBE buffer with 3 µl EtBr (10 mg/ml)

1%: 0.5 g LE agarose in a final volume of 50 ml 1X TBE buffer with 3 µl EtBr (10 mg/ml)

2%: 1 g LE agarose in a final volume of 50 ml 1X TBE buffer with 3 µl EtBr (10 mg/ml)

Agarose gel loading buffer

- 0.25% w/v bromophenol blue (0.125g)
- 40% w/v sucrose (20g)

Dissolve and make up to a final volume of 50 ml with water.

Single Stranded Conformational Polymorphism (SSCP) loading buffer (50 ml)

- 47.5 ml formamide (95%, v/v)
- 0.16 g 100 mM sodium hydroxide
- 0.125 g bromophenol blue
- 0.125 g xylene cyanol

Dissolve and make up to a final volume of 50 ml with water.

Tris Borate buffer 1X

- 125.9 g tris
- 17.3 g boric acid
- 50 µl bromophenol blue (4% w/v)

Dissolve solids and make up to a final volume of 1 litre with distilled water. Add bromophenol blue.

Tris-formate buffer (0.75M)

- 90 g tris
- Formic acid

Dissolve tris in 200 ml distilled water; adjust pH to 9.0 with formic acid and make up to a final volume of 1 litre.

40% acrylamide-PDA solution (w/v)

- 396 g acrylamide
- 4 g Piperazine diacrylamide (PDA)

Dissolve and make up to a final volume of 1 litre with distilled water.

12% non-denaturing polyacrylamide gel (v/v)

- 5.3 ml 40% (w/v) acrylamide-PDA solution
- 8.5 ml tris-formate buffer
- 3 ml 41% (v/v) glycerol
- 200 µl 10% (w/v) Ammonium Persulphate (APS)
- 20 µl TEMED

Silver staining solutions (w/v)

I. 0.1% silver nitrate solution

Dissolve solid and make up to 1 litre with distilled water.

II. Formaldehyde and sodium hydroxide solution

- 15 g sodium hydroxide pellets
- 20 ml 15% formaldehyde

Dissolve pellets first in 200 ml distilled water, add formaldehyde and make up to a final volume of 1 litre.

Luria broth (LB)

- 5 g Tryptone
- 2.5 g yeast extract
- 5 g sodium chloride

Weigh into a clean bottle. Add 500 ml distilled water and autoclave.

Selective agar plates

- 500 ml LB
- 7.5 g bacteriological agar
- 1 mg Ampicillin antibiotic dissolved in 1 ml sH_2O

Add agar to LB, autoclave and allow to cool. Once cool, add ampicillin to a final concentration of 100 μg/ml. Pour approximately 25-30 ml of agar into sterile plates (100 x 20 mm) and allow to set. Label, wrap in cling film and store at 4°C if not immediately required.

Appendix C: General protocols

Standard PCR protocol

A standard PCR experiment was performed in 0.2 ml tubes (*Merck*) as follows:

Table showing set up of a standard PCR

Reagent	Final concentration
Primers (F and R)	10 pmols
dNTPs (*Bioline*)	0.2 mM
GoTaq Reaction Buffer (*Promega*)	1X
GoTaq DNA Polymerase (*Promega*)	1 U

To each tube was added approximately 100 ng of genomic DNA and the final volume made up to 25 µl with sH_2O (*Adcock Ingram*). Thermal cycling conditions were as follows: 1 cycle of initial DNA denaturation at 95°C for 3 min; 25 cycles of (denaturation at 95°C for 30s, annealing T_a for 30s, elongation at 72°C for 40s); 1 cycle of final elongation at 72°C for 10 min. Each experiment included a negative control (containing no DNA) to test for possible contamination of reagents.

Cycle sequencing reaction

A standard cycle sequencing reaction was performed follows:

Table showing set up of cycle sequencing reaction

Reagent	Final concentration
Primers (F or R)	10 pmols
BigDye Terminator Mix (*Applied Biosystems*)	2 µl
BigDye Sequencing Buffer (*Applied Biosystems*)	1X

Each reaction contained 5-7 µl of extracted DNA or PCR product and was made up to a final volume of 20 µl with sH_2O. Thermal conditions were as follows: 1 cycle of initial DNA denaturation at 96°C for 5 min; 25 cycles of (96°C for 30s, 50°C for 15s, 60°C for 4 min); the reaction was carried out on a Perkin Elmer thermal cycler (*PerkinElmer*).

Ethanol precipitation protocol

1. Transfer cycle sequencing reaction product to a 1.5 ml eppendorf tube.
2. Add 50 µl chilled absolute ethanol and 2 µl chilled sodium acetate (appendix B).
3. Mix by vortexing and leave at -20°C overnight.
4. Centrifuge at 10000 g for 10 minutes and discard the supernatant.
5. Add 50 µl chilled 70% ethanol and mix by vortexing, repeat step 4 above.
6. Leave tube uncapped and allow to air-dry for one hour.
7. Re-suspend in 10 µl HiDi formamide (*Applied Biosystems*).

QIAquick® gel extraction kit protocol

1. Quickly and carefully excise DNA band from agarose gel using a clean sharp scalpel to minimise the amount of gel.
2. Transfer gel slice to a clean 1.5 ml eppendorf tube and weigh.
3. Add 3 volumes buffer QG to 1 volume agarose gel (100 mg ~ 100 µl).
4. Incubate tube at 50°C for 10 minutes (or until gel slice is completely melted), vortex every 2 minutes.
5. Add 1 volume isopropanol to 1 volume agarose gel and vortex.
6. Place a QIAquick spin column into the provided collection tube.
7. Apply the sample to the column and centrifuge to bind DNA.
8. Discard flow-through and replace column in the collection tube.
9. Add 0.75 ml buffer PE to the column and allow incubation at room temperature for 5 minutes. Centrifuge.
10. Discard flow-through and centrifuge again.
11. Place the column into a clean 1.5 ml tube and incubate at 37° for 5 minutes
12. Add 30 µl de-ionised water to the column and allow incubation at room temperature for 2 minutes.
13. Centrifuge to elute DNA.

All centrifuging steps in the gel extraction protocol are carried out at 13000 rpm for 1 minute, using a table top microcentrifuge (*Merck*).

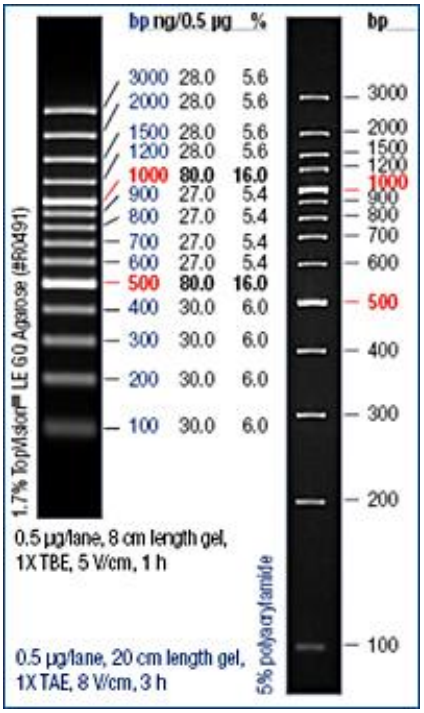
pGEM-T Easy vector ligation reaction

Table showing set up of pGEM-T Easy ligation

	Reaction (µl)	Positive control (µl)	Background control (µl)
2X Rapid ligation buffer	5	5	5
pGEM-T Easy vector	1	1	1
PCR product	X	-	-
Control insert DNA	-	2	-
T4 DNA ligase	1	1	1
sH_2O	-	1	3

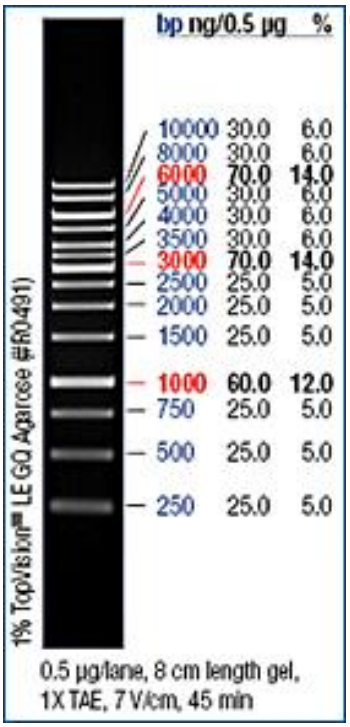
All tubes were made up to 10 µl with sH2O. The reaction tube contained X µl (i.e. up to a maximum of 3 µl) of PCR product. The positive control tube contained control insert DNA provided in the pGEM-T Easy Cloning kit. The background control tube contained no DNA to control for contamination.

Gene ruler™ 100bp Plus DNA Ladder (*Fermentas*)



(Taken from www.fermentas.com)

GeneRuler™ 1kb DNA Ladder (*Fermentas*)



(Taken from www.fermentas.com)

DNA consent form

REQUEST FOR MOLECULAR STUDIES (DNA)

Molecular Laboratory

Division of Human Genetics
IIDMM, LEVEL 3
UCT Medical School, Observatory 7925

Tel: (021) 406 6425 Fax: (021) 406 6826

Blood should be drawn in 2 plastic EDTA Tubes (Purple top) +/- 10ml each using a yellow barrel. Each tube should be inverted to mix and should be clearly labelled with the patient's name and DOB
Keep blood in fridge at 4°C until able to send to laboratory

Please DO NOT send specimens on ice or frozen.

Please fill in all the information requested:

Surname:_____ First Name(s):_____

New Family: Yes ☐ No ☐ (If no, please fill in family name) Family name: _____

Medical Aid: _____Medical Aid No: _____

Sex: M ☐ F ☐ Date of Birth: Year:_____ Month:_____ Day: _____

Number of children: _____

Ethnic Origin : (please indicate ancestry of both your mother and father)_____

Contact Address: _____ Town: _____ Fax: Tel: _____

Referring Doctor/Sister: _____ Town: _____ Fax: Tel: _____

Hospital or Address: _____ Town: _____ Fax: Tel: _____

Reason for Referral (Clinical diagnosis):

Affected ☐ At Risk ☐ Carrier ☐ Spouse ☐ Query ☐ Unaffected ☐

Becker Muscular Dys.	☐	Duchenne Muscular Dys	☐	Colonic Carcinoma	☐
Fragile-X Syndrome	☐	Bipolar Disorder	☐	Huntington Disease	☐
Retinitis Pigmentosa	☐	Spinocerebellar Ataxia	☐	Waardenberg Syndrome	☐

Additional disorders (apparent or previously treated): _____

Additional family history _____

Clinical Details:

Physical disability ☐ Mental disability ☐ Deafness ☐ Impaired vision ☐ Night blindness ☐

Other: _____

Have samples from this patient been sent to a DNA lab before? (DELETE WHERE NOT APPLICABLE) YES / NO / Don't Know

If Yes, where: _____

For Laboratory use only:

DNA number: _____ *Vol. Blood:* _____*(ml)* *Other:* _____

Date Received: Year: _____ *Month:* _____ *Day:* _____ *Computer Index No:*_____

CONSENT FOR DNA ANALYSIS AND STORAGE

1. I, _____ , request that an attempt be made using genetic material to assess the probability that: I / my child / my unborn child (DELETE WHERE NOT APPLICABLE) might have inherited a disease-causing mutation in the gene for:_____

2. I understand that the genetic material for analysis is to be obtained from: blood cells/skin sample/other (specify) (DELETE WHERE NOT APPLICABLE) :

3. I request that **no** portion of the sample be stored for later use. ☐ **(**MARK IF APPLICABLE)
   Or
   I request that a portion of the sample be stored indefinitely for (DELETE WHERE NOT APPLICABLE):
   (a) possible re-analysis
   (b) analysis for the benefit of members of my immediate family
   (c) research purposes, subject to the approval of the University of Cape Town Research Ethics Committee, provided that any information from such research will remain confidential.

4. The results of the analysis carried out on this sample of stored biological material will be made known to me, via my doctor, in accordance with the relevant protocol, if and when available.
   In addition, I authorise that they may be made known to: (DELETE WHERE NOT APPLICABLE) :
   other doctors involved in my care
   the following family members: _____
   other: _____

5. I authorise / do not authorise my doctor(s) (DELETE WHERE NOT APPLICABLE) to provide relevant clinical details to the Division of Human Genetics, UCT.

6. I have been informed that:
   (a) there are risks and benefits associated with genetic analysis and storage of biological material and these have been explained to me.
   (b) the analysis procedure is specific to the genetic condition mentioned above and cannot determine the complete genetic makeup of an individual.
   (c) the genetics laboratory is under an obligation to respect medical confidentiality .
   (d) genetic analysis may not be informative for some families or family members.
   (e) even under the best conditions, current technology of this type is not perfect and could lead to incorrect results.
   (f) where biological material is used for research purposes, there may be no direct benefit to me.

7. I understand that I may withdraw my consent for any aspect of the above at any time without this affecting my future medical care.

8. **ALL OF THE ABOVE HAS BEEN EXPLAINED TO ME IN A LANGUAGE THAT I UNDERSTAND AND MY QUESTIONS ANSWERED BY:**

_____ DATE: _____

Patient signature _____ **Witnessed consent** _____

NOTE - PLEASE INSERT A FAMILY PEDIGREE DRAWING ON THE REVERSE OF THIS FORM

Appendix D: Sequencing results

Target sequences aligned in BioEdit

Selected clone sequence aligned to wild type target A

```
-----TCGAGATATCTCAGGAGAAAAGTCCCGTGGGAAGGGTCCTGCAGGGGTGCAGGGTTGCACGC-
----
ATCGCTCGAGATATCTCAGGAGAAAAGTCCCGTGGGAAGGGTCCTGCAGGGGTGCAGGGTTGCACGCG
GCCGC
```

Selected clone sequence aligned to wild type target B

```
-----GGCCGCGTGCAACCCTGCACCCCTGCAGGACCCTTCCCACGGGACTTTTCTCCTGAGATATC-
----
CCAGCGGCCGCGTGCAACCCTGCACCCCTGCAGGGCCCTTCCCACGGGACTTTTCTCCTGAGATATCT
CGAGC
```

Selected clone sequence aligned to mutant target A

```
-----TCGAGATATCTCAGGAGAAAAGTCCCGTGGGAAGGGCCCTGCAGGGGTGCAGGGTTGCACGC-
-----
ATCGCTCGAGATATCTCAGGAGAAAAGTCCCGTGGGAAGGGCCCTGCAGGGGTGCAGGGTTGCACGCG
GCCGC
```

Selected clone sequence aligned to mutant target B

```
-----GGCCGCGTGCAACCCTGCACCCCTGCAGGGCCCTTCCCACGGGACTTTTCTCCTGAGATATC-
-----
CCAGCGGCCGCGTGCAACCCTGCACCCCTGCAGGGCCCTTCCCACGGGACTTTTCTCCTGAGATATCT
CGAGC
```

The target SNP is highlighted in red.

References

Adkins RM (2004) Comparison of the accuracy of methods of computational haplotype inference using a large empirical dataset. *BMC Genet* 5:22-29

Al-Ramahi I, Lam YC, Chen H, de Gouyon B, Zhang M, Perez AM, Branco J, de Haro M, Patterson C, Zoghbi HY and Botas J (2006) CHIP protects from the neurotoxicity of expanded and wild-type ataxin-1 and promotes their ubiquitination and degradation. *J Biol Chem* 281:26714-26724

Alves S, Nascimento-Ferreira I, Auregan G, Hassig R, Dufour N, Brouillet E, Pedroso de Lima MC, Hantraye P, Pereira de Almeida L and Déglon N (2008) Allele-specific RNA silencing of mutant ataxin-3 mediates neuroprotection in a rat model of Machado-Joseph disease. *PLoS One* 3:e3341

Alves S, Nascimento-Ferreira I, Dufour N, Hassig R, Auregan G, Nobrega C, Brouillet E, Hantraye P, Pedroso de Lima MC, Déglon N and Pereira de Almeida L (2010) Silencing ataxin-3 mitigates degeneration in a rat model of Machado-Joseph disease: no role for wild-type ataxin-3? *Hum Mol Genet* 19:2380-2394

Baer M, Nilsen TW, Costigan C and Altman S (1990) Structure and transcription of a human gene for H1 RNA, the RNA component of human RNase P. *Nucleic Acids Res* 18:97-103

Banfi S, Servadio A, Chung MY, Kwiatkowski TJ Jr, McCall AE, Duvick LA, Shen Y, Roth EJ, Orr HT and Zoghbi HY (1994) Identification and characterization of the gene causing type 1 spinocerebellar ataxia. *Nat Genet* 7:513-520

Bauer PO and Nukina N (2009) The pathogenic mechanisms of polyglutamine diseases and current therapeutic strategies. *J Neurochem* 110:1737-1765

Bonini NM and La Spada AR (2005) Silencing polyglutamine degeneration with RNAi. *Neuron* 48:715-718

Boudreau RL, Mas Monteys A and Davidson BL (2008) Minimizing variables among hairpin-based vectors reveals the potency of shRNAs. *RNA* 14:1834-1844

Brummelkamp TR, Bernards R and Agami R (2002) A system for stable expression of short interfering RNAs in mammalian cells. *Science* 296:550-552

Bryer A, Krause A, Bill P, Davids V, Bryant D, Butler J, Heckmann J, Ramesar R and Greenberg J (2003) The hereditary adult-onset ataxias in South Africa. *J Neurol Sci* 216:47-54

Burright EN, Clark HB, Servadio A, Matilla T, Feddersen RM, Yunis WS, Duvick LA, Zoghbi HY and Orr HT (1995) *SCA1* transgenic mice: a model for neurodegeneration caused by an expanded CAG trinucleotide repeat. *Cell* 82:937-948

Burright EN, Davidson JD, Duvick LA, Koshy B, Zoghbi HY and Orr HT (1997) Identification of a self-association region within the SCA1 gene product, ataxin-1. *Hum Mol Genet* 6:513-518

Castanotto D, Li H and Rossi JJ (2002) Functional siRNA expression from transfected PCR products. *RNA* 8:1454-1460

Castanotto D and Rossi JJ (2009) The promises and pitfalls of RNA-interference-based therapeutics. *Nature* 457:426-433

Chamberlain SJ, Li X and Lalande M (2008) Induced pluripotent stem (iPS) cells as *in vitro* models of human neurogenetics disorders. *Neurogenet* 9:227-235

Chen H, Fernandez-Funez, Acevedo SF, Lam YC, Kaytor MD, Fernandez MH, Aitken A, Skoulakis EMC, Orr HT, Botas J and Zoghbi HY (2003) Interaction of Akt-phosphorylated ataxin-1 with 14-3-3 mediates neurodegeneration in spinocerebellar ataxia type 1. *Cell* 113:457-468

Chen YW, Allen MD, Veprintsev DB, Lowe J and Bycroft M (2004) The structure of the AXH domain of spinocerebellar ataxin-1. *J Biol Chem* 279:3758-3765

Chong SS, McCall AE, Cota J, Subramony SH, Orr HT, Hughes MR and Zoghbi HY (1995) Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* 10:344-350

Chung MY, Ranum LP, Duvick LA, Servadio A, Zoghbi HY and Orr HT (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type 1. *Nat Genet* 5:254-258

Crespo-Baretto J, Fryer JD, Shaw CA, Orr HT and Zoghbi HY (2010) Partial loss of ataxin-1 function contributes to transcriptional dysregulation in spinocerebellar ataxia type 1 pathogenesis. *PLoS Genet* 6:e1001021

Cullen BR (2004) Derivation and function of small interfering RNAs and microRNAs. *Virus Res* 102:3-9

Cummings CJ, Mancini MA, Antalffy B, DeFranco DB, Orr HT and Zoghbi HY (1998) Chaperone suppression of aggregation and altered subcellular proteasome localization imply protein misfolding in SCA1. *Nat Genet* 19:148-154

Cummings CJ, Orr HT and Zoghbi HY (1999) Progress in pathogenesis studies of spinocerebellar ataxia type 1. *Phil Trans R Soc Lond B* 354:1079-1081

Cummings CJ, Sun Y, Opal P, Antalffy B, Mestril R, Orr HT, Dillmann WH and Zoghbi HY (2001) Over-expression of inducible HSP70 chaperone suppresses neuropathology and improves motor function in SCA mice. *Hum Mol Genet* 14:1511-1518

Cvetanovic M, Rooney RJ, Garcia JJ, Toporovskaya N, Zoghbi HY and Opal P (2007) The role of LANP and ataxin 1 in E4F-mediated transcriptional repression. *EMBO Rep* 7:671-677

Davies SW, Turmaine M, Cozens BA, DiFiglia M, Sharp AH, Ross CA, Scherzinger E, Wanker EE, Mangiarini L and Bates GP (1997) Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* 90:537-548

Davies SW, Beardsall K, Turmaine M, DiFiglia M, Aronin N and Bates GP (1998) Are neuronal intranuclear inclusions the common neuropathology of triplet-repeat disorders with polyglutamine-repeat expansions? *Lancet* 351:131-133

de Chiara C, Menon RP, Strom M, Gibson TJ and Pastore A (2009) Phosphorylation of S776 and 14-3-3 binding modulate ataxin-1 interaction with splicing factors. *PLoS One 4*:e8372

de Wit E, Delport W, Rugamika CE, Meintjes A, Möller M, van Helden PD, Seoighe C and Hoal EG (2010) Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. *Hum Genet* 128:145-153

Ding H, Schwarz DS, Keene A, Affar EB, Fenton L, Xia X, Shi Y, Zamore PD and Xu Z (2003) Selective silencing by RNAi of a dominant allele that causes amyotrophic lateral sclerosis. *Aging Cell* 2:209-217

Dorschner MO, Barden D and Stephens K (2002) Diagnosis of five spinocerebellar ataxia disorders by multiplex amplification and capillary electrophoresis. *J Mol Diagn* 2:108-113

Drouet V, Perrin V, Hassig R, Dufour N, Auregan G, Alves S, Bonvento G, Brouilett E, Luthi-Carter R, Hantraye P and Déglon N (2009) Sustained effects of nonallele-specific *huntingtin* silencing. *Ann Neurol* 65:276-285

Ebert AD, Yu J, Rose Jr FF, Mattis VB, Lorson CL, Thomson JA and Svendsen CN (2009) Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* 457:277-280

Emamian ES, Kaytor MD, Duvick LA, Zu T, Tousey SK, Zoghbi HY, Clark HB and Orr HT (2003) Serine 776 of ataxin-1 is critical for polyglutamine-induced disease in *SCA1* transgenic mice. *Neuron* 38:375-387

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806-811

Gonzalez-Alegre P, Bode N, Davidson BL, and Paulson HL (2005) Silencing primary dystonia: lentiviral-mediated RNA interference therapy for DYT1 dystonia. *J Neurosci* 45:10502-10509

Gonzalez-Alegre P and Paulson HL (2007) Technology insight: therapeutic RNA interference - how far from the neurology clinic? *Nature* 7:394-404

Goold R, Hubank M, Hunt A, Holton J, Menon RP, Revesz T, Pandolfo M and Matilla-Dueñas A (2007) Down-regulation of the dopamine receptor D2 in mice lacking ataxin 1. *Hum Mol Genet* 17:2122-2134

Greenberg J, Solomon GAE, Vorster AA, Heckmann J and Bryer A (2006) Origin of the SCA7 mutation in South Africa: implications for molecular diagnostics. *Clin Genet* 70:415-417

Grimm D, Streetz KL, Jopling CL, Storm TA, Pandey K, Davis CR, Marion P, Salazar F and Kay MA (2006) Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* 441:537-541

Grimm D and Kay MA (2009) RNAi and gene therapy: a mutual attraction. *Hematology Am Soc Hematol Educ Program* 2007:473-481

Hammond SM (2005) Dicing and slicing, the core machinery of the RNA interference pathway. *FEBS Lett* 579:5822-5829

Harper SQ, Staber PD, He X, Eliason SL, Martins IH, Mao Q, Yang L, Kotin RM, Paulson HL and Davidson BL (2005) RNA interference improves motor and neuropathological abnormalities in a Huntington's disease mouse model. *Proc Natl Acad Sci USA* 102:5820-5825

Hong S, Lee S, Cho S and Kang S (2008) UbcH6 interacts with and ubiquitinates the SCA1 gene product ataxin-1. *Biochem Biophys Res Commun* 371:256-260

Jodice C, Malaspina P, Persichetti F, Novelletto A, Spadaro M, Giunti P, Morocutti C, Terrenato L, Harding AE and Frontali M (1994) Effect of trinucleotide repeat length and parental sex on phenotypic variation in spinocerebellar ataxia 1. *Am J Hum Genet* 54:959-965

Kaytor MD and Warren ST (1999) Aberrant protein deposition and neurological disease. *J Biol Chem* 274:37507-37510

Kennedy L and Shelbourne PF (2000) Dramatic mutation instability in HD mouse striatum: does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease? *Hum Mol Genet* 9:2539-2544

Kim DH and Rossi JJ (2007) Strategies for silencing human disease using RNA interference. *Nat Rev Genet* 8:173-184

Klement IA, Skinner PJ, Kaytor MD, Yi H, Hersch SM, Clark HB, Zoghbi HY and Orr HT (1998) Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice. *Cell* 95:41-53

Krol HA, Krawczyk PM, Bosch KS, Aten JA, Hol EM and Reits EA (2008) Polyglutamine expansion accelerates the dynamics of ataxin-1 and does not result in aggregate formation. *PLoS One* 3:e1503

La Spada AR and Taylor JP (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat Rev Genet* 11:247-258

Lam YC, Bowman AB, Jafar-Nejad P, Lim J, Richman R, Fryer JD, Hyun ED, Duvick LA, Orr HT, Botas J and Zoghbi HY (2006) ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. *Cell* 127:1335-1347

Lee Y, Jeon K, Lee J, Kim S and Kim VN (2002*a*) MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* 21:4663-4670

Lee NS, Dohjima T, Bauer G, Li H, Li MJ, Ehsani A, Salvaterra P and Rossi J (2002*b*) Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nat Biotechnol* 20:500-505

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S and Kim VN (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425:415-419

Lee S, Hong S and Kang S (2008) The ubiquitin-conjugating enzyme UbcH6 regulates the transcriptional repression activity of the SCA1 gene product ataxin-1. *Biochem Biophys Res Commun* 372:735-740

Leuschner PJF, Ameres SL, Kueng S and Martinez J (2006) Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep* 7:314-320

Li Y, Yokota T, Matsumura R, Taira K and Mizusawa H (2004) Sequence-dependent and independent inhibition specific for mutant ataxin-3 by small interfering RNA. *Ann Neurol* 56:124-129

Lim J, Crespo-Barreto J, Jafar-Nejad P, Bowman AB, Richman R, Hill DE, Orr HT and Zoghbi HY (2008) Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1. *Nature* 452:713-719

Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song J, Hammond SM, Joshua-Torr L and Hannon G (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305:1437-1441

Lorenzetti D, Watase K, Xu B, Matzuk MM, Orr HT and Zoghbi HY (2000) Repeat instability and motor incoordination in mice with a targeted expanded CAG repeat in the *Sca1* locus. *Hum Mol Genet* 9:779-785

Marchetto MCN, Winner B and Gage FH (2010) Pluripotent stem cells in neurodegenerative and neurodevelopmental diseases. *Hum Mol Genet* 19:R71-R76

Margolis RL (2002) The spinocerebellar ataxias: order emerges from chaos. *Curr Neurol Neurosci Rep* 2:447-456

Matilla T, Volpini V, Genis D, Rosell J, Corral J, Davalos A, Molins A and Estivill X (1993) Presymptomatic analysis of spinocerebellar ataxia type 1 (SCA1) via the expansion of the SCA1 CAG-repeat in a large pedigree displaying anticipation and parental male bias. *Hum Mol Genet* 2:2123-2128

Matilla A, Roberson ED, Banfi S, Morales J, Armstrong DL, Burright EN, Orr HT, Sweatt JD, Zoghbi HY, and Matzuk MM (1998) Mice lacking ataxin-1 display learning deficits and decreased hippocampal paired-pulse facilitation. *J Neurosci* 18:5508-5516

McBride JL, Boudreau RL, Harper SQ, Staber PD, Mas Monteys A, Martins I, Gilmore BL, Burstein H, Peluso RW, Polisky B, Carter BJ and Davidson BL (2008) Artificial miRNAs mitigate shRNA-mediated toxicity in the brain: implications for the therapeutic development of RNAi. *Proc Natl Acad Sci USA* 105:5868-5873

Meister G and Tuschl T (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature* 431:343-349

Miller VM, Xia H, Marrs GL, Gouvion CM, Lee G, Davidson BL and Paulson HL (2003) Allele-specific silencing of dominant disease genes. *Proc Natl Acad Sci USA* 100:7195-7200

Mittal U, Sharma S, Chopra R, Dheeraj K, Pal PK, Srivastava AK and Mukerji M (2005) Insights into the mutational history and prevalence of SCA1 in the Indian population through anchored polymorphisms. *Hum Genet* 118:107-114

Mizutani A, Wang L, Rajan H, Vig PJS, Alaynick WA, Thaler JP and Tsai C (2005) Boat, an AXH domain protein, suppresses the cytotoxicity of mutant ataxin-1. *EMBO J* 24:3339-3351

Moolman-Smook JC, De Lange WJ, Bruwer ECD, Brink PA and Corfield V (1998) The origins of hypertrophic cardiomyopathy-causing mutations in two South African sub-populations: a unique profile of both independent and founder events. *Am J Hum Genet* 65:1308-1320

Ohnishi Y, Tamura Y, Yoshida M, Tokunaga K and Hohjoh H (2008) Enhancement of allele discrimination by introduction of nucleotide mismatches into siRNA in allele-specific gene silencing by RNAi. *PLoS One* 3:e2248

Okazawa H, Rich T, Chang A, Lin X, Waragai M, Kajikawa M, Enokido Y, Komuro A, Kato S, Shibata M, Hatanaka H, Mouradian MM, Sudol M and Kanazawa I (2002) Interaction between mutant ataxin-1 and PQBP-1 affects transcription and cell death. *Neuron* 34:701-713

Orr HT, Chung M, Banfi S, Kwiatkowski TJ Jr, Servadio A, Beaudet AL, McCall AE, Duvick LA, Ranum LP, and Zoghbi HY (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* 4:221-226

Orr HT and Zoghbi HY (2001) SCA1 molecular genetics: a history of a 13 year collaboration against glutamines. *Hum Mol Genet* 10:2307-2311

Parfitt DA, Michael GJ, Vermeulen EGM, Prodromou NV, Webb TR, Gallo J, Cheetham ME, Nicoll WS, Blatch GL and Chapple JP (2009) The ataxia protein sacsin is a functional co-chaperone that protects against polyglutamine-expanded ataxin-1. *Hum Mol Genet* 18:1556-1565

Park Y, Hong S, Kim SJ, and Kang S (2005) Proteasome function is inhibited by polyglutamine-expanded ataxin-1, the SCA1 gene product. *Mol Cells* 19:23-30

Perez MK, Paulson HL, Pendse SJ, Saionz SJ, Bonini NM and Pittman RN (1998) Recruitment and the role of nuclear localization in polyglutamine-mediated aggregation. *J Cell Biol* 143:1457-1470

Quintana-Murci L, Haramant C, Quach H, Balanovsky O, Zaporozhchenko V, Bormans C, van Helden PD, Hoal EG and Behar DM (2010) Strong maternal Khoisan contribution to the South African Coloured Population: A case of gender-biased admixture. *Am J Hum Genet* 86:611-620

Ramesar RS, Bardien S, Beighton P and Bryer A (1997) Expanded CAG repeats in spinocerebellar ataxia (SCA1) segregate with distinct haplotypes in South African families. *Hum Genet* 100:131-137

Riley BE, Zoghbi HY and Orr HT (2005) SUMOylation of the polyglutamine repeat protein, ataxin-1, is dependent on a functional nuclear localisation signal. *J Biol Chem* 280:21942-21948

Rodriguez-Lebron E and Gonzalez-Alegre P (2006) Silencing neurodegenerative disease: bringing RNA interference to the clinic. *Expert Rev Neurother* 6:223-233

Rodriguez-Lebron E and Paulson HL (2006) Allele-specific RNA interference for neurological disease. *Gene Ther* 13:576-581

Rohlfs RV and Weir BS (2008) Distributions of Hardy-Weinberg equilibrium test statistics. *Genetics* 180:1609-1616

Ross CA (1997) Intranuclear neuronal inclusions: a common pathogenic mechanism for glutamine-repeat neurodegenerative diseases? *Neuron* 19:1147-1150

Scholefield J and Greenberg J (2007) A common SNP haplotype provides molecular proof of a founder effect of Huntington disease linking two South African populations. *Eur J Hum Genet* 15:590-595

Scholefield J, Greenberg LJ, Weinberg MS, Arbuthnot PB, Abdelgany A and Wood MJA (2009) Design of RNAi hairpins for mutation-specific silencing of ataxin-7 and correction of a SCA7 phenotype. *PLoS One* 4:e7232

Scholefield J and Wood MJA (2010) Therapeutic gene silencing strategies for polyglutamine disorders. *Trends Genet* 26:29-38

Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N and Zamore PD (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115:199-208

Schwarz DS, Ding H, Kennington L, Moore JT, Schelter J, Burchard J, Linsley PS, Aronin N, Xu Z and Zamore PD (2006) Designing siRNA that distinguish between genes that differ by a single nucleotide. *PLoS Genet* 2:e140

Servadio A, Koshy B, Armstrong D, Antalffy B, Orr HT and Zoghbi HY (1995) Expression analysis of the ataxin-1 protein in tissues from normal and spinocerebellar ataxia type 1 individuals. *Nat Genet* 10:94-98

Shao Y, Chan C, Maliyekkel A, Lawrence CE, Roninson IB and Ding Y (2007) Effect of target secondary structure on RNAi efficiency. *RNA* 13:1631-1640

Sheth U and Parker R (2003) Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* 300:805-808

Sibley CR, Seow Y and Wood MJA (2010) Novel RNA-based strategies for therapeutic gene silencing. *Mol Ther* 18:466-476

Singh SK and Hajeri PB (2009) siRNAs: their potential as therapeutic agents-Part II. Methods of delivery. *Drug Discov Today* 14:859-865

Skinner PJ, Koshy BT, Cummings CJ, Klement IA, Helin K, Servado A, Zoghbi HY and Orr HT (1997) Ataxin-1 with an expanded glutamine tract alters nuclear matrix-associated structures. *Nature* 389:971-974

Soong B and Paulson HL (2007) Spinocerebellar ataxias: an update. *Curr Opin Neurol* 20:438-446

Starfield M, Hennies HC, Jung M, Jenkins T, Wienker T, Hull P, Spurdle A, Kuster W, Ramsay M and Reis A (1997) Localization of the gene causing keratolytic winter erythema to chromosome 8p22-p23, and evidence for a founder Effect in South African Afrikaans-speakers. *Am J Hum Genet* 61:370-378

Stephens M and Donelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73:1162-1169

Timchenko LT and Caskey CT (1999) Triplet repeat disorders: discussion of molecular mechanisms. *Cell Mol Life Sci* 55:1432-1447

Tokui K, Adachi H, Waza M, Katsuno M, Minamiyama M, Doi H, Tanaka K, Hamakazi J, Murata S, Tanaka F and Sobue G (2009) 17-DMAG ameliorates polyglutamine-mediated motor neuron degeneration through well-preserved proteasome function in an SBMA model mouse. *Hum Mol Genet* 18:898-910

Tsai C, Kao H, Mitzutani A, Banayo E, Rajan H, McKeown M and Evans RM (2004) Ataxin 1, a SCA1 neurodegenerative disorder protein, is functionally linked to the silencing mediator of retinoid and thyroid hormone receptors. *Proc Natl Acad Sci USA* 101:4047-4052

Tsuda H, Jafar-Nejad H, Patel AJ, Sun Y, Chen H, Rose MF, Venken KJT, Botas J, Orr HT, Bellen HJ and Zoghbi HY (2005) The AXH domain of ataxin-1 mediates neurodegeneration through its interaction with Gfi-1/Senseless proteins. *Cell* 122:633-644

Umahara T and Uchihara T (2010) 14-3-3 proteins and spinocerebellar ataxia type 1: from molecular interaction to human neuropathology. *Cerebellum* 9:183-189

Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Südhof TC and Wernig M (2010) Direct conversion of fibroblast to functional neurons by defined factors. *Nature* 463:1035-1041

Watase K, Weeber EJ, Xu B, Antalffy B, Yuva-Paylor L, Hashimoto K, Kano M, Atkinson R, Sun Y, Armstrong DL, Sweatt JD, Orr HT, Paylor R and Zoghbi HY (2002) A long CAG repeat in the mouse Sca1 locus replicates SCA1 features and reveals the impact of protein solubility on selective neurodegeneration. *Neuron* 34:905-919

Watase K, Venken KJT, Sun Y, Orr HT and Zoghbi HY (2003) Regional differences of somatic CAG repeat instability do not account for selective neuronal vulnerability in a knock-in mouse model of SCA1. *Hum Mol Genet* 12:2789-2795

Weinberg MS and Wood MJA (2009) Short non-coding RNA biology and neurodegenerative disorders: novel disease targets and therapeutics. *Hum Mol Genet* 18:R27-R39

Wigginton JE, Cutler DJ and Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet* 76:887-893

Wood MJA, Trülzsch B, Abdelgany A and Beeson D (2003) Therapeutic gene silencing in the nervous system. *Hum Mol Genet* 2:R279-R284

Xia H, Mao Q, Eliason SL, Harper SQ, Martins IH, Orr HT, Paulson HL, Yang L, Kotin RM and Davidson BL (2004) RNAi suppresses polyglutamine-induced neurodegeneration in a model of spinocerebellar ataxia. *Nat Med* 10:816-820

Yamada M, Sato T, Tsuji S, and Takahashi H (2008) CAG repeat disorder models and human neuropathology: similarities and differences. *Acta Neuropathol* 115:71-86

Yi R, Qin Y, Macara IG and Cullen BR (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* 17:3011-3016

Yue S, Serra HG, Zoghbi HY and Orr HT (2001) The spinocerebellar ataxia type 1 protein, ataxin-1, has RNA-binding activity that is inversely affected by the length of its polyglutamine tract. *Hum Mol Genet* 10:25-30

Zhang Y, Engelman and Friedlander RM (2009) Allele-specific silencing of mutant Huntington's disease gene. *J Neuorochem* 108:82-90

Zoghbi HY, Jodice C, Sandkuijl LA, Kwiatkowski TJ, McCall AE, Huntoon SA, Lulli P, Spadaro M, Litt M, Cann HM, Frontali M and Terrenato L (1991) The gene for autosomal dominant spinocerebellar ataxia (SCA1) maps telomeric to the HLA complex and is closely linked to the D6S89 locus in three large kindreds. *Am J Hum Genet* 49: 23-30

Zoghbi HY (2000) Spinocerebellar ataxias. *Neurobiol Dis* 7:523-527

Zoghbi HY and Orr HT (2000) Glutamine repeats and neurodegeneration. *Annu Rev Neurosci* 23:217-247

Zoghbi HY and Orr HT (2009) Pathogenic mechanisms of a polyglutamine-mediated neurodegenerative disease, spinocerebellar ataxia type 1. *J Biol Chem* 284:7425-7429

Zu T, Duvick LA, Kaytor MD, Berlinger MS, Zoghbi HY, Clark HB and Orr HT (2004) Recovery from polyglutamine-induced neurodegeneration in conditional *SCA1* transgenic mice. *J Neurosci* 24:8853-8861

Zühlke C, Dalski A, Hellenbroich Y, Bubel S, Schwinger E and Bürk K (2002) Spinocerebellar ataxia type 1 (SCA1): phenotype-genotype correlation studies in intermediate alleles. *Eur J Hum Genet* 10:204-209

Electronic resources

- Basic Local Alignment Search Tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi)
- Bioedit sequence alignment editor (v7.05.2) Hall, T.A. (1999)

  Bioedit: a user friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucl Acids Symp Ser. 41:95-98*
- Ensembl genome browser (http://www.ensembl.org)
- GeneMapper Software (v3.0, *Applied Biosystems*)
- Hardy-Weinberg Equilibrium

  (http://ihg2.helmholtz-muenchen.de/cgi-bin/hw/hwa1.pl)
- National Centre for Biotechnology Information (NCBI)

  (http://www.ncbi.nlm.nih.gov)
- OligoAnalyzer (v3.0, C 2007 Integrated DNA Technologies)
- OligoCalculator (Buehler)
- PHASE (v2.1, Stephens *et al.* 2001)

  http://www.stat.washington.edu/stephens/software
- Primer3 software (v.0.4.0) S. Rosen and HJ Skaletsky 2000
- Webcutter 2.0 (C) 1997 Max Heiman